

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2002, 14:29:45 ; Search time 57.98 Seconds
(without alignments)
8922.114 Million cell updates/sec

Title: US-09-856-327-1
Perfect score: 2106
Sequence: 1 atcagcatgtctctctca.....aaaaaaaaaaaaaaaaaaaaa 2106

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents,NA:*
1: /cgn2_6/pdata/1/ina/5A_COMB.seq:*
2: /cgn2_6/pdata/1/ina/5B_COMB.seq:*
3: /cgn2_6/pdata/1/ina/6A_COMB.seq:*
4: /cgn2_6/pdata/1/ina/6B_COMB.seq:*
5: /cgn2_6/pdata/1/ina/PCUS_COMB.seq:*
6: /cgn2_6/pdata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	212.4	10.1	1869	3	US-09-305-381-1
2	210.2	10.0	1869	1	US-08-734-925-1
3	152.6	7.2	1701	4	US-09-023-731-2
4	88.8	4.2	1733	3	US-09-073-569-1
5	87	4.1	2239	4	US-09-196-390-1
6	86.6	4.1	2671	6	5168051-9
7	85	4.0	1582	3	US-08-545-196B-10
8	85	4.0	1582	3	US-08-545-196B-12
9	84.4	4.0	1641	1	US-08-300-903A-8
10	84.4	4.0	3410	4	US-09-020-956-110
11	84.4	4.0	3410	4	US-09-030-607-110
12	84.4	4.0	3410	4	US-09-439-313-110
13	83.6	4.0	1474	4	US-08-821-994-64
14	83.4	4.0	1602	1	US-08-530-950-3
15	83.4	4.0	1602	3	US-08-888-429A-3
16	83.4	4.0	1602	4	US-09-149-879-3
17	83.2	4.0	2852	3	US-09-027-137-2
18	83	3.9	1198	3	US-09-248-335-35
19	83	3.9	1134	3	US-09-248-335-29
20	83	3.9	1193	4	US-09-372-422A-23
21	82.6	3.9	1813	4	US-09-071-224-3
22	82.4	3.9	144	1	US-08-702-344-26
23	82.4	3.9	1117	4	US-09-247-373B-33
24	82.4	3.9	1307	2	US-08-960-022-17
25	82.4	3.9	2082	2	US-08-785-310A-2
26	82.2	3.9	1736	3	US-09-182-816-22
27	82.2	3.9	1736	3	US-09-182-816-24

C

ALIGNMENTS

RESULT 1

US-09-305-381-1
; Sequence 1, Application US/09305381
; Patent No. 6145865
; GENERAL INFORMATION:
; APPLICANT: Christensen, Soren
; APPLICANT: Lassen, Soren Flensted
; APPLICANT: Schneider, Falle
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; TITLE OF INVENTION: Having Pyranose Oxidase Activity
; FILE REFERENCE: 5571.200-US
; CURRENT APPLICATION NUMBER: US/09/305,381
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/088,724
; EARLIER FILING DATE: 1998-06-10
; EARLIER APPLICATION NUMBER: PA 1998 00774
; EARLIER FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1869
; TYPE: DNA
; ORGANISM: Trametes hirsuta
; US-09-305-381-1

Query Match 10.1%; Score 212.4; DB 3; Length 1869;
Best Local Similarity 51.5%; Pred. No. 6.5e-41;
Matches 841; Conservative 0; Mismatches 656; Indels 135; Gaps 10;

QY	269	gggtacacaaagaagaatgaatcgagttccagaaagattatgacgcgtctcgcaatgta	328
DB	259	gggtacacaaagaagaacacccgtcgagtaaccagaagaacatcgacaaatcgtaattgt	318
QY	329	atcaaggagccttacaacaaagtctctgttcctgcagaaaccagacgctgcacatt	388
DB	319	atcaaggagcacttatgcccgtctcggtgcccgtcaacacgatggtcgctgacacgcta	378
QY	389	gatccggagcctggagcgcgccttcggaagttcagccatctcgaacggtataaaatcct	448
DB	379	agcccggctcattggaagcttcgacg-----ttcttcgcgaacggcggaatcca	432
QY	449	cacacagcgggaattcagaaacttgctgcggagcgttaacgcgtgagtcgcggcgtatg	508
DB	433	gagcaagacccgtcgcaaccttagtggccagcgggtcaccgcgtcgtcggcgcatg	492
QY	509	agtaccactggagtgctccacgcacggtattcaccacccatggaaagtctcccggtc	568
DB	493	tctacgcactggagtgctcgacgcgcgtcttcgagaagctgcag-----	537

Query Match 10.0%; Score 210.2; DB 1; Length 1869;
Best Local Similarity 50.4%; Pred. No. 2.1e-40;
Matches 857; Conservative 0; Mismatches 713; Indels 129; Gaps 9;

QY 200 gqatcgagagctctatagcttctacgctgttaatgcgaagaagaactgcagttccc 259
DB 190 GAGCTCTCGAAGCGGTTCAGAGTGCCTCATGTTGACATCGGGAAATGACTCTGGC 249

QY 260 tacgttctggtaccacaagaagaatgaaatcgagttccagaagaatattgaccgttc 319
DB 250 CTGAAGATCGTGCCACAGAGAAGAACACCGTCGAATACCAAGAACATGACAACTTT 309

QY 320 gtcaatgtaatcaaggagagccttacaacaagtctctgttctcttcagaaacagaacgtg 379
DB 310 GTGAACGTCAATCAGGGCCCAATGATGTCGTTCCTCCGTCATATACCCCTCGTGATC 369

QY 380 cctacacttgatcccgagcgtgagcgcccccgtggaagtccagcatatgaaaggt 439
DB 370 GACACGCTCAGCCGAGCTTTGGCAAGCTTCATCG-----TCTTTCGTCGCGCAATGGC 423

QY 440 aaaaactctcaccagcggaattcgaagaacttctgctgagggccgttaacgctggagtc 499
DB 424 TCGAACCCAGCAGGAGCCGCTTCGTAACTCAGTGGTCAGGGGTCACGCGTGTGTC 483

QY 500 ggcggcatgtagtaccactgagcgtgctccacgcgcagcgattcatccaccatggaaagt 559
DB 484 GGAGGCATGTCCACGCACACTGGCATGGCGCACACCGCGCTTGA----- 527

QY 560 ctccgggcatcgccgtcgaagctcagtaacgacccggcagagcagaagaagtgg 619
DB 528 --CGCGAGCAGCGCCCGTTCGTCTGTAAGAGCAGCACCAGACGCTGACGCGCGATGG 585

QY 620 aacgagctttatccagggcgagcgtctcatcgggacttccaccagggaattcgaag 679
DB 586 GACCGCTGTACACCAAGCGGAGTCATCTTCAAGACCGGAGCGGACCACTTCAAGGAG 645

QY 680 tcaattggcaacccttctgctgctcttcttgaagacgctgataaggtatcgtaacgt 739
DB 646 TCGATCGCCACCAACCTCGTGTCTCAACAAGCTCGCGGAGGAATACAAAGTCAAGCGCGAC 705

QY 740 attcttgcctctccgttggatcgccacggttgaaagacgcgccggaatcgctgaa 799
DB 706 TTCAGCAGATCCGCTCGCGCGCAACCGCTCG-----CAGTCCGACCTTCGTGCGAG 756

QY 800 tggactcagcagaagaatctttccactctatctacacgatgacaagcagaagaagctc 859
DB 757 TGGAGCTCGGCGAACACCGGTGT-----CGACCTCCAGAACAGCGCGAAC 801

QY 860 ttatccctgtgacgaaccatcgctgcacacgactggcgttacggcggttatgagaag 919
DB 802 ACGGACGCGCGGAATGAGCGCTTCAACCTCTTCCCGCGGTTCATGTGAGCGCTGCTG 861

QY 920 aagattggcgtcgaggtcagaatctactgtgcccacaggaatcctagtctgcagctg 979
DB 862 CGCAACACGTGCAATCCGAGATCGGAGAGTCTGCACATCCACACCTTCATCTCGGCGAC 921

QY 980 gacagctatatcgccgaagggtatagtactggcgtcgaggcgtatcggaaccacag 1039
DB 922 CGCTTCGAATCAAGCAGACGCTGTCGTTCATACAGCCGGGGGTCCACACGCGCAG 981

QY 1040 attctctataactgggctt-----ctctgggtacaggttcacgcgaacgaatgactcg 1093
DB 982 CTCTCTGTGAACCTCTGGCTTTGGACAGCTGGCGCGCGGACCCCGGAACCGCGCAG 1041

QY 1094 ttgatcccaactgggaggtacatcacgagcagccagtcattttccagatagtc 1153
DB 1042 TTCTGTCGCTTCTGGGAAGCTACATCCAGCAGCGTCTGCTCTCTCTGCCAGACCGTG 1101

QY 1154 ttgaggcaggaattcgtcgcagcgtg-----cgcgacgactctatggactgcc----- 1203
DB 1102 ATGAGCACCAGCTCATCGACCGGTCAAGTCCGACATGATCATCAGGGCGCAACCTTGGC 1161

QY 1204 -----atgg 1207

DB 1162 GATCTGGGTACAGCGTACGTACACCGCCGCGGAGACCAACAAGCACCCGAGCTGG 1221

QY 1208 tggaagaagccgtgtgtcaacatatgtccaagaaccgcagatgcactgccccattccg 1267

DB 1222 TGAACGAAAGGTGAAGAACCATATGTCAGCACCCAGGAGACCGCTTCCAAATCCCG 1281

QY 1268 ttccgcgatccggaacccccaggtaaacacccccatttaccagaagaacccccctggcacag 1327

DB 1282 TTCGAGGACCCCGAGCGCAGGTACACACCTGTGTTCCAGCCATCGCACCGTGGCACACT 1341

QY 1328 cagattccacgcgatgcttttttctgtaacggtgcgctgctgaggtggaactctctgtc 1387

DB 1342 CAGATTACCGCGATGCGTTTTCAGTTAGCGGCGGTGAGCAAAAGCATCGACTCACGCTC 1401

QY 1388 atcgtgacctgcgtgtgttgcgaacccgacccctgaagaacaaacacctttgtgtttc 1447

DB 1402 ATGCTGACCTGCGCTTCTTCGCGCGGACGAGGCAAGGAGGAGAAACAAAGCTCTGGTTC 1461

QY 1448 cagaacgatgtcaagacgggtacagttatgccgcagccgacttccagatatcgaccacgc 1507

DB 1462 TCGACAAAAATTACGGACACGTACACATGCCGACGCGAGCTTCGACTTCCGGCTCCCG 1521

QY 1508 actg---cgtcaacgtgagagcaagaaatgatggccgatatgtgcgaagtggcgagc 1564

DB 1522 GCGGGCGCACGACGAAAGGAGGAGGAGCATGATGACCGATATGTGCTTATGTCGGCG 1581

QY 1565 aactggagagttatttgccacgctcccccgcagtttatgattcagacgcttgcactt 1624

DB 1582 AAGATTGCTGGTTCCTGCGCGCGCTCCCTCCGCAATTATGAGAGCGCGCTTGTCTCTT 1641

QY 1625 catcttgggggactactcgcattggttctgcac-----aaggcaactacagttggtgat 1678

DB 1642 CACTCGTGTGTAGCGCACCGCATGGGCTTCGACGAGCAGGAGGACAAAGTGTGCGTCAAC 1701

QY 1679 acaactcgtgtgtgtggactttgccaattctttatgttcagggcaatggcaccatcagg 1738

DB 1702 ACGGATCGGCGCTGTTTGGCTTCAAGAACCTGTTCCTCGTGGCTCGGAAACATTTCCC 1761

QY 1739 acgggctcgcgagaaaccgcacacttgcagtgtgccacgtatcaagagcgcgagg 1798

DB 1762 ACGCGTACGCGCGGAACCCGACCTCACCGAATGTGCGTCCGATCAAGAGTTCGCGAG 1821

QY 1799 agcatcatcaatacactca 1817

DB 1822 TACATCAAGAAACAATTCA 1840

RESULT 3
US-09-023-731-2
; Sequence 2, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akhiero; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731

[illegible]

4

US-09-196-390-1
; Sequence 1, Application US/09196390
; Patent No. 6307125
; GENERAL INFORMATION:
; APPLICANT: Block, Martina
; APPLICANT: Lorz, Horst
; APPLICANT: Lutticke, Stephanie
; APPLICANT: Walter, Lennart
; APPLICANT: Froberg, Claus
; APPLICANT: Kossmann, Jens
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING ENZYMES
; TITLE OF INVENTION: FROM WHEAT WHICH ARE INVOLVED IN STARCH
; TITLE OF INVENTION: SYNTHESIS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James F. Haley, Jr., c/o Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,390
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 21 588.9
; FILING DATE: 29-MAY-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 36 917.7
; FILING DATE: 11-SEP-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP97/02793
; FILING DATE: 28-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley, Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: AGREVO-9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Triticum aestivum L.
; STRAIN: cv. Florida
; HAPLOTYPE: ca. 21 d Caryopses
; IMMEDIATE SOURCE:
; LIBRARY: cDNA library in pBluescript sk (-)
; CLONE: TASSS
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..2017
US-09-196-390-1

Query Match 4.1%; Score 87; DB 4; Length 2239;
Best Local Similarity 67.2%; Pred. No. 2.1e-11;
Matches 123; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
Qy 1924 tctgtccctaccatgtgatgtacgatagcgttgaaagattttgtgtattactga 1983

Db 2054 TGTGTCTGTGCTTAGCTGACAAATATTTGACCTGTTGGAGAAATTTTATCTTTGCTGCT 2113
Qy 1984 acctgtacttctgaatgattgacacatgatcatgttttaaaaaa 2043
Db 2114 GTTTTATTAATCAAAAGAGGGGTTCCCTCCGATTTCATTA 2173
Qy 2044 aaaaaa 2103
Db 2174 aaaaaa 2233
Qy 2104 aaa 2106
Db 2234 AAA 2236
RESULT 6
5168051-9
; Patent No. 5168051
; APPLICANT: DERYNCK, RIK M.A.; GOEDEL, DAVID V.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING TGF-B ITS USES
; NUMBER OF SEQUENCES: 21
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/389,929
; FILING DATE: 04-AUG-1989
; SEQ ID NO: 9
; LENGTH: 2671
5168051-9
Query Match 4.1%; Score 86.6; DB 6; Length 2671;
Best Local Similarity 83.8%; Pred. No. 2.7e-11;
Matches 98; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
Qy 1990 acttctgtgaatgattgacacatgatcatgttttaaaaaa 2049
Db 2544 atgttcttaacataaagccttattcgaagtgttaaaaaa 2603
Qy 2050 aaaaaa 2106
Db 2604 aaaaaa 2660
RESULT 7
US-08-545-196B-10
; Sequence 10, Application US/08545196B
; Patent No. 6080577
; GENERAL INFORMATION:
; APPLICANT: MELKI, JUDITH
; APPLICANT: MUNNICH, ARNOLD
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
; TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,196B
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. J.
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 2121-110P


```
Db 1628 AAAAAA 1633
|||||
RESULT 10
US-09-020-956-110
; Sequence 110, Application US/09020956
; Patent No. 6261562
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,956
; FILING DATE: 09-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
;
US-09-020-956-110

Query Match 4.0%; Score 84.4; DB 4; Length 3410;
Best Local Similarity 80.8%; Pred. No. 9.9e-11;
Matches 97; Conservative 1; Mismatches 22; Indels 0; Gaps 0;

QY 1987 tgtactttgtcgaatagtcagtcactatgattcattgttataaaataaaataaaataaa 2046
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3287 TGTATTGTTGACAAATTAAGGCTTTCTATATGTTTAAAAAATAAATAAATAAATAA 3346
QY 2047 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2106
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3347 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAATAA 3406

RESULT 11
US-09-030-607-110
; Sequence 110, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
```

```
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3410 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
;
US-09-030-607-110

Query Match 4.0%; Score 84.4; DB 4; Length 3410;
Best Local Similarity 80.8%; Pred. No. 9.9e-11;
Matches 97; Conservative 1; Mismatches 22; Indels 0; Gaps 0;

QY 1987 tgtactttgtcgaatagtcagtcactatgattcattgttataaaataaaataaa 2046
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3287 TGTATTGTTGACAAATTAAGGCTTTCTATATGTTTAAAAAATAAATAAATAAATAA 3346
QY 2047 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2106
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 3347 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAATAA 3406

RESULT 12
US-09-439-313-110
; Sequence 110, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqi
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 110
; LENGTH: 3410
; TYPE: DNA
; ORGANISM: Homo sapien
;
US-09-439-313-110
```

```
Query Match          4.0%; Score 84.4; DB 4; Length 3410;
Best Local Similarity 80.8%; Pred. No. 9.9e-11;
Matches 97; Conservative 1; Mismatches 22; Indels 0; Gaps 0;

Qy 1987 tgtacttgctgaatgattgacacatgattcattgtttataaaaaaa 2046
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3287 tggttatggtacaaaataaaggcttcttattgtttataaaaaaa 3346

Qy 2047 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2106
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3347 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 3406

RESULT 13
US-08-821-994-64
; Sequence 64, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 64
; LENGTH: 1474
; TYPE: DNA
; ORGANISM: Brassica napus
US-08-821-994-64

Query Match          4.0%; Score 83.6; DB 4; Length 1474;
Best Local Similarity 75.4%; Pred. No. 1.1e-10;
Matches 104; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 1969 ttgtgtattactgaacctgtactttgtctgaatgattggtcactatgattcattgta 2028
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1333 ttatgtattaagagagtataataaatagatattcttcttaaaaaaa 1392

Qy 2029 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2088
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1393 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1452

Qy 2089 aaaaaaaaaaaaaaaaaa 2106
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1453 aaaaaaaaaaaaaaaaaa 1470

RESULT 14
US-08-530-950-3
; Sequence 3, Application US/08530950
; Patent No. 5736381
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Raingeaud, Joel
; APPLICANT: Gupta, Shashi
; APPLICANT: Derijard, Benoit
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND ONCOPROTEIN-
; TITLE OF INVENTION: ONCOPROTEIN-ACTIVATED HUMAN PROTEIN KINASE
; TITLE OF INVENTION: KINASES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
```

```
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,950
; FILING DATE: 19-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/010001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1602 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-530-950-3

Query Match          4.0%; Score 83.4; DB 1; Length 1602;
Best Local Similarity 93.5%; Pred. No. 1.3e-10;
Matches 87; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2014 tatgattcatgtttataaaaaaa 2106
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1509 TATGAGTCCTCAAAAAA 1601

Qy 2074 aaaaaaaaaaaaaaaaaa 2106
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1569 AAAAAA 1601

RESULT 15
US-08-888-429A-3
; Sequence 3, Application US/08888429A
; Patent No. 6136596
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Whitmarsh, Alan
; APPLICANT: Tournier, Cathy
; TITLE OF INVENTION: CYTOKINE-, STRESS-, AND ONCOPROTEIN-
; TITLE OF INVENTION: ACTIVATED HUMAN PROTEIN KINASE
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/888,429A
; FILING DATE: 07-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/530,950
; FILING DATE: 19-SEP-1995
; APPLICATION NUMBER: 08/446,083
```

; FILING DATE: 19-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, Peter J.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/053001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 299354
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1602 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 244...1245
; US-08-888-429A-3

Query Match 4.0%; Score 83.4; DB 3; Length 1602;
Best Local Similarity 93.5%; Pred. No. 1.3e-10;
Matches 87; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 2014 tatgattcatgtttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2073
||||| ||| ||||||||||||||||||||||||||||||||||||||||
Db 1509 TATGAGTCCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1568
Qy 2074 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2106
||||||||||||||||||||||||||||||||||||||
Db 1569 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1601

Search completed: September 3, 2002, 15:51:01
Job time: 4876 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2002, 14:36:36 ; Search time 242.59 Seconds
(without alignments)
14905.074 Million cell updates/sec

Title: US-09-856-327-1
Perfect score: 2106
Sequence: 1 atcagcatgtctctctca.....aaaaaaaaaaaaaaaaaaaaa 2106

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2106	100.0	2106	22	AAF99980 Nucleotide sequenc
2	212.4	10.1	1869	22	AAF99980 Trametes hirsuta p
3	212.4	10.1	1935	22	AAF99980 Trametes hirsuta p
4	208.6	9.9	1869	17	AAF34420 Pyranose oxidase e
5	175.6	8.3	1902	21	AAZ46411 Pleurotus cornucop
6	152.6	7.2	1701	20	AAV83626 Nucleic acid encod
7	129	6.1	1946	21	AAV71487 T. matsutake pyran
8	129	6.1	1946	21	AAV71487 Trichoderma derive
9	94.4	4.5	4639	22	AA545962 Human DNA encoding

10	94.4	4.5	4640	21	AA78483 Human PRO708 (UNO3
11	94.4	4.5	4650	20	AA333987 Human PRO708 nucle
12	91.8	4.4	2440	22	AAH34932 Human colon cancer
13	91.6	4.3	2323	19	AAV59524 Human secreted pro
14	91.4	4.3	1480	22	AD07771 Human secreted pro
15	91.2	4.3	2710	21	AA54132 Breast cancer prot
16	90.8	4.3	1327	24	AD24775 Glycine max ankyri
17	90.6	4.3	1392	22	AAF72748 Human prostate can
18	89.6	4.3	374	22	AAI91055 Human polynucleoti
19	89.6	4.3	1046	24	AA63134 Cell death protect
20	89.6	4.3	2260	22	AAI97914 Human neuroblastom
21	89.6	4.3	2260	22	AAI98068 Human neuroblastom
22	89.4	4.2	1735	22	AA503904 Human breast cance
23	89.2	4.2	1735	22	AA503904 Human secreted pro
24	88.8	4.2	1733	20	AAV81394 Human tumour antig
25	88.4	4.2	400	22	AAI84784 Human polynucleoti
26	88.4	4.2	1091	22	AA89723 Maize ZmGnsN1-1 gl
27	88.2	4.2	421	22	AAI91775 Human polynucleoti
28	88	4.2	422	22	AAI87735 Human polynucleoti
29	88	4.2	425	22	AA60450 Human cancer agent
30	88	4.2	567	22	AAH33915 Human colon cancer
31	88	4.2	1248	21	AA26436 Human secreted pro
32	87.8	4.2	310	22	AAH71505 Human cervical can
33	87.8	4.2	823	22	AAO7664 Human ovarian and
34	87.8	4.2	823	22	AAO7664 Human reproductive
35	87.8	4.2	3312	8	AA71034 pmx34 cDNA insert.
36	87.6	4.2	393	22	AAI11189 Human breast cance
37	87.6	4.2	10039	24	ABL34042 Human immune syste
38	87.4	4.2	545	22	AAH70126 Human cervical can
39	87.4	4.2	546	22	AAH71167 Human cervical can
40	87.4	4.2	1199	20	AAO8695 Novel nucleotide s
41	87.2	4.1	765	21	AA59300 Human secreted pro
42	87.2	4.1	1254	22	AA525886 Human cDNA encodin
43	87.2	4.1	2434	21	AA26402 Human secreted pro
44	87	4.1	379	22	AAI20476 Human breast cance
45	87	4.1	392	22	AAI88688 Human polynucleoti

ALIGNMENTS

RESULT 1
AAF99980 standard; cDNA; 2106 BP.
XX AAF99980;
AC AAF99980;
XX 20-JUL-2001 (first entry)

XX Nucleotide sequence encoding Lyophyllum shimeji antibacterial protein.
DE Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;
KW Pyricularia oryzae; Rhizoctonia solani; rice pathogen; ss.
XX Lyophyllum shimeji.
XX Key Location/Qualifiers
FH 8..1864
CDS /*tag- a
FT /product= "antibacterial protein"

XX WO200121657-A1.
XX 29-MAR-2001.
XX 20-SEP-2000; 2000WO-JP06404.
XX 21-SEP-1999; 99JP-0267238.
XX (NISB) JAPAN TOBACCO INC.
XX (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
XX Takakura Y, Kuwata S, Inoue Y;

Db 1741 gggcttcggcgagaccgacacattacgtcgtatgtgccacgctatcaagagcgcgaggag 1800

QY 1801 catcataatcacactcaagggtggactgacggagaaataacacagcgagcatcgcaacct 1860

Db 1801 catcataatcacactcaagggtggactgacggagaaataacacagcgagcatcgcaacct 1860

QY 1861 ttgaggaagagcaacagcagtgtaaacaaacgctcaagtggtcactctcaagttaagt 1920

Db 1861 ttgaggaagagcaacagcagtgtaaacaaacgctcaagtggtcactctcaagttaagt 1920

QY 1921 cattctggctccctaccattgtatgtgtacgataggcgttgaagatttggattac 1980

Db 1921 cattctggctccctaccattgtatgtgtacgataggcgttgaagatttggattac 1980

QY 1981 tgaacctgtactttgtctgaatgtatgtgacactatgattcatgttttaaaaaaaaaa 2040

Db 1981 tgaacctgtactttgtctgaatgtatgtgacactatgattcatgttttaaaaaaaaaa 2040

QY 2041 aa 2100

Db 2041 aa 2100

QY 2101 aaaaaa 2106

Db 2101 aaaaaa 2106

RESULT 2

AAC87519

ID AAC87519 standard; DNA; 1869 BP.

XX AC AAC87519;

XX DT 13-MAR-2001 (first entry)

XX DE Trametes hirsuta pyranose oxidase cDNA, SEQ ID NO:1 (version 2).

XX KW Pyranose oxidase; expression construct; recombinant production;

XX KW monosaccharide oxidation; 2-keto derivative;

XX KW hydrogen peroxide production; ss.

OS Trametes hirsuta.

XX US6146865-A.

PD 14-NOV-2000.

PF 05-MAY-1999; 99US-0305381.

XX 08-JUN-1998; 98DK-0000774.

PR 10-JUN-1998; 98US-0088724.

XX (NOVO) NOVO NORDISK AS.

XX PI Schneider P, Christensen S, Lassen SF;

XX WPI; 2001-049055/06.

DR P-PSDB; AAB48832.

XX Novel nucleic acid molecule encoding polypeptide having pyranose

PT oxidase activity used to design oligonucleotide probes to identify and

PT clone DNA encoding the polypeptide from different genera or species -

XX Claim 2; Column 25-28; 20pp; English.

XX The invention relates to nucleic acids (e.g., AAC87518, AAC87519) which

CC encode Trametes hirsuta pyranose oxidase (AAB48832). The invention also

CC relates to expression constructs, expression vectors and recombinant

CC cells comprising pyranose oxidase nucleic acid sequences, and the

CC recombinant production of Trametes hirsuta pyranose oxidase. Pyranose

CC oxidase catalyses the oxidation of several monosaccharides in the

CC pyranose form at position C2 to produce 2-keto derivatives with the

CC release of hydrogen peroxide. Nucleic acids encoding Trametes hirsuta

CC pyranose oxidase may be used to produce the enzyme and to design

CC oligonucleotide probes to identify and clone genomic pyranose oxidase

CC cDNA or genomic DNA from different genera or species of microorganisms

CC (fungi or bacteria). The present sequence represents a cDNA encoding

CC pyranose oxidase from the fungus Trametes hirsuta.

CC Note: Both AAC87518 and AAC87519 are Trametes hirsuta pyranose oxidase

CC cDNA sequences which contain the entire open reading frame (ORF).

CC However, the two sequences have different stop codons - AAC87518 has an

CC opal stop codon while AAC87519 has an amber stop codon.

XX SQ Sequence 1869 BP; 397 A; 622 C; 548 G; 302 T; 0 other;

Query Match 10.1%; Score 212.4; DB 22; Length 1869;

Best Local Similarity 51.5%; Pred. No. 2.2e-25;

Matches 841; Conservative 0; Mismatches 656; Indels 135; Gaps 10;

QY 269 ggtacacacagaagaatgaaatcgattgtccagaagaatattaccgcttcgcaatgta 328

Db 259 ggtcacacaagaagacacccgtcgagtaccagaagaacatcgacaattcgttaattgt 318

QY 329 atcaaggagcccttacaacaagtctctgttctctgcagaaaccagacgtgcctacatt 388

Db 319 atacaagggaacttatgccgtctcgtggtccggtcaacacgatggtcgttgacagcta 378

QY 389 gatccggagcctggagcgcccccttgaaagtccagccatcgcgaacggtaaaatcct 448

Db 379 agccggcgctatggcaagcttcgacg-----ttctgtccgcaacggggcgaaatcca 432

QY 449 caccagcgggaattcagaaactgtctgaggagcgtgaacgctgagtcgcggcagtcg 508

Db 433 gagcaagaccgctgcgcacaccttagtgcagcggcgtcaccgcgctcgtcgcggcagtc 492

QY 509 agtaccactggagctgctccacgcacacgattcattccacccttggaagtctcccgggc 568

Db 493 tctacactggagctgctgcgacgcgcttcgagagctgcag-----537

QY 569 atcgccgttcggaagctcagtaacacccggcgagagcagacaaagtggaacagctt 628

Db 538 ---cgcccgtcgtcgtggaagaacgactccaaggcgagcagcggagtggaacagctc 594

QY 629 tattcagagcggagcgtctcattcgggaacttcacacaaaggaattcagcagtgcaattcgg 688

Db 595 tacaagaagcggagctgctactcaagacggcgacacccagcttcgcgagtgatcgcg 654

QY 689 cacaccttgtctgcgctctttgcaagacgcgtcaacaggtatcgtaacgtatcttcgcg 748

Db 655 cacaacctcgtgctcaagaagctgcaggaggagtagtaaaagg---gcgtgcgacttcag 711

QY 749 cctctcccgctggcatgccaccggttgaagaacgcgccggaatacgtcgaaatggcactca 808

Db 712 cagatcccgctcgcggcgacgcgccaga-----gccgacgcttcgtcagtgagactcg 765

QY 809 gcagaaaattctttccactctatctacaacgtagcaagcagagaagaagctcttaccctg 868

Db 766 gcgcacacgctgtt-----cgatctcgagaacccggccgaacaggagcgcg 810

QY 869 ctgacaaacatcgctgcacacgactggcgttacggcggttatgagaagaagattggc 928

Db 811 ccgaagcagcgttcaacctcttcccgctcgtgcgtgcacgacgtgagcgcgataac 870

QY 929 gctgcgaggtcaggaatctactgtccacccaggaaatcctagtctcgagctggacagctat 988

Db 871 gcgaactcggagatcgtaggccttgatgtccgcacctccacgggggcaagacatcacc 930

QY 989 ataatggcaaggtatattatgtactgctggcgagcgtggaatcggaacccacagattctctat 1048

Db 931 atcaagccaaaggtgtacatccctcacccgcgcggttcacacacgcgcagctccctcgcg 990

QY 1049 aactcggggtt-----ctctgggtacaggtcagccacgcgaatgac---tcgttgatc 1099

Db 991 gcctctggattcggcgagctgggtcgtcccgaccgccacccgcaagcgcgtgcgtctctgct 1050

```
Qy 1100 cccaactctggaggagtaacatcacgcagcagccgattggtccagatagtttgagg 1159
Db 1051 ccgtacctggggaccacatcacgcagcagcgtctctctgacagacgctcatgagc 1110
Qy 1160 caggaattcgtcgacagcagcgtcgagacatc----- 1190
Db 1111 acggagctcatcaacagtgctcacgcgagatagaccattgtcggaagcccgccacccg 1170
Qy 1191 -----cttatggactgccatggtggaaa 1213
Db 1171 gactatagctcagctatacccccgggcaacccgaacaacaagcaccgcgaactggtggaa 1230
Qy 1214 gaagcgtgtgtcaacatttgcaagaacccgcagacagatgcactgcccattccgttcgcg 1273
Db 1231 gagaaggtgaagaagcacatgatgaccacagagagagccgcgtcccgatcccggtcgag 1290
Qy 1274 gatccggaacccaggttaaacacccatttacagaagaacacccctggcgcacacgcagatt 1333
Db 1291 gaccctgagccgaggtcaccacgctgtttcaggcaacgcccattggcacacaccagatt 1350
Qy 1334 caccgcatgcttttctgacggtgcccgtgctgaggtgagctcgtgctatcgtc 1393
Db 1351 caccgcagccttcagctacgagcgcgtgcagcagacatgcactgcggtcatcgtc 1410
Qy 1394 gactcgcgtgtgttggcgcaacgcagccctgaagcaaaacaccttttggttttccagaac 1453
Db 1411 gactggtgttcttcggcgcagcagccagcccaagagaggaagcagatggtttctcggac 1470
Qy 1454 gatgtcaagcgggtacagtatccgcagcgcagcttcagatatcgaccacgactgcg 1513
Db 1471 aagatcacggagcgtacaacctccgagcgcgttcgacttcctgcttcccccggg-- 1528
Qy 1514 tcaaacgtgagagcaagaaatgatgctgcatatgtgcgaagtgcgagcaacttggga 1573
Db 1529 ---gccgcgaagcggagacatgatgaccacagatgctgcatgctgcggaagatcggt 1584
Qy 1574 ggttattgcccacgtcccccccgagtttatgatccagcgttcgacttcattcttgcg 1633
Db 1585 ggaattcgtcgtggtcttaaccacagttcatgagcccggttcttctgctgcaaccttgg 1644
Qy 1634 gggactactcgtattggttgcacaa-----ggcaactacagtggtgtgatacaaacctcg 1687
Db 1645 gggacgcaccgcagtggttcgagcagagagcgaagcgdaagtgctgctgcgaccgcactca 1704
Qy 1688 ctggtctgggactttgccaactctttatgttcaggcgaatggcaccatcagagcggcttc 1747
Db 1705 cgcgtcttcggttcaagaacaccttctcctgcgcggtcgcgggaacatccccaccgcgtac 1764
Qy 1748 ggcgagaacccgacacttacgtgatgtgcccacgctatcaagagcgcgagagcatcatc 1807
Db 1765 gccgcgaacccgagcgtcacgcgaatgctgcttgcatcaagagctgctgagatcatcaag 1824
Qy 1808 aatacacatcaag 1819
Db 1825 aagaacttcgag 1836
```

RESULT 3

AAC87518

ID AAC87518 standard; DNA; 1995 BP.

XX

AC AAC87518;

XX

DT 13-MAR-2001 (first entry)

XX

DE Trametes hirsuta pyranose oxidase cDNA, SEQ ID NO:1 (version 1).

XX

KW Pyranose oxidase; expression construct; recombinant production;

KW monosaccharide oxidation; 2-keto derivative;

KW hydrogen peroxide production; ss.

XX

OS Trametes hirsuta.

XX

PN US6146865-A.

XX

PD 14-NOV-2000.

XX

PF 05-MAY-1999; 99US-0305381.

XX

PR 08-JUN-1998; 98DK-0000774.

PR 10-JUN-1998; 98US-0088724.

XX

PA (NOVO) NOVO NORDISK AS.

XX

PI Schneider P, Christensen S, Lassen SF;

XX

DR WPI; 2001-049055/06.

XX

DR P-PSDB; AAB48832.

XX

Novel nucleic acid molecule encoding polypeptide having pyranose oxidase activity used to design oligonucleotide probes to identify and clone DNA encoding the polypeptide from different genera or species - Claim 2; Fig 1; 20pp; English.

The invention relates to nucleic acids (e.g., AAC87518, AAC87519) which encode Trametes hirsuta pyranose oxidase (AAB48832). The invention also relates to expression constructs, expression vectors and recombinant cells comprising pyranose oxidase nucleic acid sequences, and the recombinant production of Trametes hirsuta pyranose oxidase. Pyranose oxidase catalyses the oxidation of several monosaccharides in the pyranose form at position C2 to produce 2-keto derivatives with the release of hydrogen peroxide. Nucleic acids encoding Trametes hirsuta pyranose oxidase may be used to produce the enzyme and to design oligonucleotide probes to identify and clone genomic pyranose oxidase cDNA or genomic DNA from different genera or species of microorganisms (fungi or bacteria). The present sequence represents a cDNA encoding pyranose oxidase from the fungus Trametes hirsuta.

Note: Both AAC87518 and AAC87519 are Trametes hirsuta pyranose oxidase cDNA sequences which contain the entire open reading frame (ORF).

However, the two sequences have different stop codons - AAC87518 has an opal stop codon while AAC87519 has an amber stop codon.

Sequence 1995 BP; 437 A; 649 C; 574 G; 335 T; 0 other;

Query Match 10.1%; Score 212.4; DB 22; Length 1995;

Best Local Similarity 51.5%; Pred. No. 2.2e-25;

Matches 841; Conservative 0; Mismatches 656; Indels 135; Gaps 10;

Qy 269 ggctaccacaagaagaatgaaatcgagttccagaagaatattgaccgcttcgtcaatgta 328

Db 288 ggctcacacaagaagaacacgctcgagttaccagaagaacatcgacaaattgtaaatgt 347

Qy 329 atcaaggagccttacacaagtctcttctctcagaaacaccagaacgctccacatt 388

Db 348 atacaagggaactatgcccgtctcggtccgcctcgaacacgattgctggaacgcta 407

Qy 389 gatccggagcctggagcgcgccttcggaagtgcacatcatcgaaacgctaaatact 448

Db 408 agcccgctcatggcaagcttcgag-----ttcttcgcccgaacggcggaatcca 461

Qy 449 caccagcgggaattcgagaacttctctcgaggccgtaacgctgagtggtcgcgcatg 508

Db 462 gagcaagaccgcgtgcgcaacttagtggtgcagggcgtcaccccgctcgtcgcgcatg 521

Qy 509 agtaccactgagcgtctccacccacgattatccaccatggaagatccccgggc 568

Db 522 tctcagcactggagcgtgcgacccgcgcttcgagaagctgcag----- 566

Qy 569 atcggcgctcggaagctcagtaacgacccgcgacgagcagcagaagagtggaagcgtt 628

Db 567 --cgcccgctcgtcgtaagaagactcccaagcgcgacgacgcgagtggaagcgtc 623

Qy 629 tattccgagggccgagcgtctctatcggtggaactccaccagggaattcgacgagtcattcg 688

```
Db 624 tacaagaagcgagtcgtacttcaagacagcgccacgacccagttccgagtcgacccg 683
Qy 689 cacaccctgttctgcgtctcttgcagacgcgtacaagatgcgtacaagatctcttcgc 748
Db 684 cacaacctgcttcaagaagctgcagagagagtacaagg---gcgtgcgacttccag 740
Qy 749 cctctcccggttgctacgcacacggttgaagaagcgcccggaatacgtgaatggcactca 808
Db 741 cagatcccgctcgcgacgcgcagcaga-----gcccgacgtctcagtgagtgctcg 794
Qy 809 gcagaaatcttttccactctatctacaacgatgacaagcagaagaagctctttaccctcg 868
Db 795 gcgcacacgctgt-----cgatctcgagaacccgcccgaagaagcgacgcg 839
Qy 869 ctgacgaacatcgctgcacacgactggcgcttacggtggcggtgatgagaagaagattggc 928
Db 840 ccgaagcagcgcttcaacctctcccgcgctgcgtgcacgaacgtgagcgcgataac 899
Qy 929 gctgcgaggtcaaggaatctacttgccacacaggaatccttagttcgacgtcgacagctat 988
Db 900 gggaaactcgagatcgtaggcctggatgtccgcgacctccacggggggaagagatcacc 959
Qy 989 atcatggcaagatgatgtactggcgtgcggagcgatcggaacccacagattctctat 1048
Db 960 atcaaggcaaggtgtacatctcaacccgcgcgcgtccacacgcgcagctccctcg 1019
Qy 1049 aactcgggtt-----ctctgggttacaggtcacgcccacgcaatgac--tcgttgatc 1099
Db 1020 gctctggattcggcagctgggtctccgcaccccgccaagcgcgtgcgtctctgcgtg 1079
Qy 1100 ccaactctgggaggtacatcacggagcagcgcgatggcattttgcagatagctcttgagg 1159
Db 1080 cgttactggggaccacatcacccagcagagcgtcgtcttcttgccagacgctcatgagc 1139
Qy 1160 caggaaatctgcagacagctggcgacgac----- 1190
Db 1140 acggagctatcaacagtgctaccgcggatagacattgtgcggcaagcccgccaccgcg 1199
Qy 1191 -----cttatggactgcacatggtgga 1213
Db 1200 gactatagcgtcacgtatacccgggcaacccgaaacaagcaccgcgactggtggaac 1259
Qy 1214 gaagcgtgtgctacaacatattccaaagaacccgacagatgcactgccattccgtccgc 1273
Db 1260 gagaagtgaaagaagcacatgatggaccaccagagaccgcgtcccgatcccgcttcgag 1319
Qy 1274 gatccggaacccaggtaaacaacccatttacagaagaacacccctggcacacagagatt 1333
Db 1320 gaccttgagccgaggtcaccacgctgtttcagggaacgcacccatggcacaccccgatt 1379
Qy 1334 caccgcatgcttttcgtacggtgcgctcggtcctgaggtggaactcgtgtcatcgtc 1393
Db 1380 caccgcgacgcttcagctacgcgcgcgtgcagcagacatgcgactcggctcatcgtc 1439
Qy 1394 gactcgcgtgtgttgccgaacccgacctgaagcaaaacacacttttggttttccagaac 1453
Db 1440 gactgcggttcttcgacgcacccgacccaaagaggaagaacagatattggttctcgac 1499
Qy 1454 gatgttcaagagcgggtacagatgctgcgacgcgacgttcagatatcgaccacgactcgc 1513
Db 1500 aagatcacgagcgcgtacaaacctccggcgagcgacgttcgactccgctcccccggg-- 1557
Qy 1514 tcaacgctgagagcaggaagaaatgatggcgcgtatgtcggaagtggcgagcaacttggga 1573
Db 1558 ----gccgcgaagcggagacatgatgacgcacatgtcgtcatgtcgcggaagatcgtg 1613
Qy 1574 ggttatttccacagctcccccgcaggttatggatccaggcgttcgacttcatcttgcg 1633
Db 1614 ggattcgtcgtgggtctccaccacagttcatggagcccggtcttctcgtcacttgg 1673
Qy 1634 gggactactcagattgcttcgacaa-----ggcaactacagtggtgctatacaactcg 1687
Db 1674 gggacgcacgcgcatgggtctcagcagaaggggacaaagtgctgctcagaccgactca 1733
```

```
Qy 1688 ctggctcgggacttcccaatctttatgttcagggaatggcaccatcaggcgggcttc 1747
Db 1734 cgcgtctcggcttcaagaacctcttccctcggcgctgcgggaacatccccccgctac 1793
Qy 1748 ggcgagaacccgacacattacgtcgtatgtccacgctatcaagagcgcggagagcatc 1807
Db 1794 gcgcgaaacccgacgtcccgcaatgtcgttcgacgaagctgcgagtcacatcaag 1853
Qy 1808 aatacactcaag 1819
Db 1854 aagaacttcgag 1865
```

RESULT 4

AAT34420
ID AAT34420 standard; cDNA; 1869 BP.

XX AAT34420;

XX 27-NOV-1996 (first entry)

XX Pyranose oxidase encoding sequence.

DE Pyranose oxidase; glucose; oxidation; glucanase; assay; diabetes;
KW marker; diagnosis; 1,5-anhydro-D-sorbitol; ss.

XX Coriolus versicolor.

XX DE19545780-A1.

XX 13-JUN-1996.

XX 07-DEC-1995; 95DE-1045780.

XX 24-MAY-1995; 95JP-0124835.

PR 07-DEC-1994; 94JP-0304086.

XX (KIRK) KIRKMAN CORP.

XX Kawai G, Koyama Y, Minamihara T, Nishimura I, Okada K,
PI Suzuki M;

XX WPI; 1996-278990/29.

DR P-PSDB; AAR99628.

XX DNA encoding protein with pyranose oxidase activity at neutral pH -
PT useful for the determination of glucose in body fluids or foods, or
PT 1,5-anhydro-D-sorbitol used as marker for diabetes diagnosis

XX Claim 1; Page 10-13; 22pp; German.

XX The present sequence encodes a protein isolated from Coriolus versicolor,
CC which has the enzyme activity of pyranose oxidase (PO). The PO oxidises
CC glucose to gluconone and has an optimum pH of 7-7.5. It has a mol. wt. of
CC 290000 (determined by gel filtration) and is stable at around 50deg.C.
CC PO can be used for measurement of glucose in, e.g. foods or body fluids,
CC or 1,5-anhydro-D-sorbitol which is an important marker used in the
CC diagnosis of diabetes.

XX Sequence 1869 BP; 408 A; 601 C; 521 G; 339 T; 0 other;

Query Match

Best Local Similarity 9.9%; Score 208.6; DB 17; Length 1869;
Matches 856; Conservative 0; Mismatches 714; Indels 129; Gaps 9;

```
Qy 200 gagatcgagctgctgatagcttctacgctgttaatgccgaagaagaactcagttccc 259
Db 190 gagctcgtgaagcgggttacaaaggtcgcacgttcgtacatcggggaattgactctggc 249
Qy 260 tacgttctcgttcaccacaagaagaatgaatcgagttccacagaagatatgacgcttc 319
```

Db	250	ctgaagatcgggtgccacaagaagaacacccgctcgaataccagaagaacattgacaagttt	309
Qy	320	gtcaattgaatacaaggagagccttacacaagctctctgtctctgtcagaacaacgaacgtg	379
Db	310	gtgaacgtcaattcagggccaattgatgtctgttccgttcccgccaataccctcgtgac	369
Qy	380	cctacacttgatcccgaggagccctggagcgcgcccccttgaagtccagccatatcgaaacggt	439
Db	370	gaacgctcagccgcgcgtctgtggaagttcatcg-----ttcttcgtccgcaatggcg	423
Qy	440	aaaaatcctaccagcgggaattcgagaacttgctcgggagggccgttaacgctggagatc	499
Db	424	tcgaaccagagcaggaaccccgcttcgttaactccagtgttcagggcggtcaacgctgtcgt	483
Qy	500	ggcgcatgagtaccacatggagctgtccacgcgcacggattcatccaccatggaaagt	559
Db	484	ggaggcatgtccacgcacgtggacatgctcgacacgcgcgtttga-----atgg	527
Qy	560	ctccgggcatcgccgcgtccgaagctcagtaacgaccgcggcagagcgacgaagaagatgg	619
Db	528	--ccgcgagcagcgcccgctgtcgtggaaggagccacgagcgctgacgacgcgcagtg	585
Qy	620	aacgagctttattccgaggccgagcgtctcattcggaacttccaccaaggaattcgacgag	679
Db	586	gaccgctgttacccaaggccgagtcatacttcaagaccgcggaacgagttccaagag	645
Qy	680	tcaattcggcacacctgtctcgcgtttgtcgaagcgcgtacaaggatcgtcaacgt	739
Db	646	tcgatacgcacaaacctcgtgtccaaagctcgcgggggaatacaaaaggccagcgac	705
Qy	740	atctttgcctctcccgcttggtcagtcgaccgcgtttgaagaacgcgcggaatacgtcgaa	799
Db	706	ttccagcagatccctcgcgcgcaacgcgctcg-----cagtcgcgacctcgtcgag	756
Qy	800	tggcactcagcagaataatttttccactctatctacaacgatgacaagcagaagaagctc	859
Db	757	tggagctcggcgaaacccgtgtt-----cgacctccagaaagcgcgaac	801
Qy	860	ttaccctcgtcagcaacctcgtgcacacgactggccttacggctacggcggtatgagaag	919
Db	802	acggacgcgcgcgaatgagcgttcaacctcttcccgcgcgttgcatgtgagcgcgtcgtg	861
Qy	920	aagattggcgtccgaggttcaggaatactactggccaccaggaatcctagtctcgacgtg	979
Db	862	cgcaaacgctcgaaactccgagatcgagagctgcacatccacgacctcatctccgcgcac	921
Qy	980	gacagctatatacgcgaaggtatatgtactggcgtcgggagcgatcggcaacccacag	1039
Db	922	cgcctcgaaatacaaacagacgtgttcgttctatacgcgcgggcgtccacaacgcgcag	981
Qy	1040	attctataactcggcctt-----ctcgggctacaggttcagcgcacgcaatgactcg	1093
Db	982	cttctcgtgaaactctggtcttggaacgtcgggcgcgcgcgcgcgcgaaccccgccgag	1041
Qy	1094	ttgatcccaacactggggaggtacatacgcgagcgcgcgatggcaatttgcagatagtc	1153
Db	1042	ttgctgcgcctcgcgaagctacatcacgcgagcagtcgctcgtcttctgcgaacgcgtg	1101
Qy	1154	ttaggcaggaataatgctcagacgcgtg-----cgcgacgataccttatgactgcgc-----	1203
Db	1102	atgagcacgcgagctatcagacgcgtcaagtcgcgacatgatcatcagggggcaacctggc	1161
Qy	1204	-----atgg	1207
Db	1162	gatactggggtacagcgttcacgttacacgcgcgcgcgcgcgcgcgaacaaacgaagca	1221
Qy	1208	tggaaagaagccgtttgctcaacatatgtccaaagaaccgcgacagatgcaactgccattccg	1267
Db	1222	tggaaacgaagaagtgaagaacacacatgatgcagacacaggaggaccgcgttccaatcccg	1281
Qy	1268	ttccgcgataccggaaaccccgaggtatacaacccccattatcagaagaacacccccctggcaacg	1327
Db	1282	ttcagagaccccagcgcgaggtccaccacttcttccagccatcgtcagccctgtgcacact	1341

RESULT	5	
AAZ46411		
ID	AAZ46411	standard; DNA; 1902 BP.
XX	AC	
XX	AAZ46411;	
XX	07-MAR-2000	(first entry)
XX	Pleurotus cornuopieae	antitumour protein coding sequence.
DE		
XX	Antitumour;	cancer; tumour; treatment; expression; tumour suppressor;
KW	p53;	pBR; ss.
XX		
OS	Pleurotus cornuopieae.	
XX		
PN	JPI1315096-A.	
XX		
PD	16-NOV-1999.	
XX		
PF	07-AUG-1998;	98JP-0236349.
XX		
PR	08-AUG-1997;	97JP-0215311.
PR	02-MAR-1998;	98JP-0066176.
XX		
PA	(NEWF-) NEW FOOD CREATION GIJUTSU KENKYU KUMIAI.	
XX		
DR	WPI; 2000-058170/05.	
DR	P-PSDB; AAY52700.	
XX		
PT	An antitumour protein derived from Pleurotus cornuopieae and its gene	
PT	useful for treatment of cancer including those caused by abnormal	
PT	expression of cancer inhibitory gene (e.g. p53 and pBR)	
XX		
PS	Claim 9; Page 15-16; 23pp; Japanese.	
XX		

PR 11-APR-2000; 2000US-196187P.
PR 11-APR-2000; 2000US-196690P.
PR 11-APR-2000; 2000US-196820P.
PR 18-APR-2000; 2000US-198121P.
PR 18-APR-2000; 2000US-198585P.
PR 25-APR-2000; 2000US-199397P.
PR 25-APR-2000; 2000US-199505P.
PR 25-APR-2000; 2000US-199654P.
PR 03-MAY-2000; 2000US-201516P.
PR 17-MAY-2000; 2000WO-US13705.
PR 22-MAY-2000; 2000WO-US14042.
PR 30-MAY-2000; 2000WO-US14941.
PR 02-JUN-2000; 2000WO-US15264.
PR 05-JUN-2000; 2000US-209832P.
PR 28-JUL-2000; 2000WO-US20710.
PR 22-AUG-2000; 2000US-0644848.
PR 24-AUG-2000; 2000WO-US23328.
PR 08-NOV-2000; 2000WO-US30952.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000WO-US34956.
XX
PA (GETH) GENENTECH INC.
XX
XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
XX
DR WPI: 2001-602746/68.
DR P-PSDB; AAU29061.
XX
PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT presence of tumours, such as prostate and breast tumours, in mammals and
PT to screen for modulators of the compounds -
XX
PS Claim 2; Fig 75; 774pp; English.
XX
CC Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
CC primers for PRO polypeptides of the invention. The sequences of the
CC invention can be used to detect the presence of a tumour in a mammal by
CC comparing the level of expression of a PRO polypeptide in a test sample
CC of cells from the animal and a control sample of normal cells, whereby a
CC higher level of expression in the test sample indicates the presence of a
CC tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
CC pigs, goats and rabbits but are preferably human. The polypeptides can be
CC used to stimulate tumour necrosis factor (TNF) alpha release from human
CC blood, when contacted with it. A specific polypeptide can be used to
CC stimulate the proliferation or differentiation of chondrocyte cells. The
CC PRO proteins can be used to determine the presence of tumours and also
CC susceptibility to tumour development, particularly adrenal, lung, colon,
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC can be used for genetic analysis of individuals with genetic disorders.
XX
SQ Sequence 4639 BP; 1426 A; 956 C; 1025 G; 1232 T; 0 other;

Query Match 4.5%; Score 94.4; DB 22; Length 4639;
Best Local Similarity 72.9%; Pred. No. 1.2e-06;
Matches 137; Conservative 0; Mismatches 46; Indels 5; Gaps 1;

Qy 1924 tctgtccctaccatgtgtgtatgtagcagtagcggttgaaagatttgggtactactga 1983
Db 4423 tctgtccctaccatgtgtgtatgtagcagtagcggttgaaagatttgggtactactga 4482
Qy 1984 acctgtactttgtctga----atagttatgacctatgattcatgtttataaaaaaa 2038
Db 4483 cagtatcttttaaaagagctcttaaaaaaaagggcatattgttcagttaaaaaaa 4542
Qy 2039 aaaaaaa 2106
Db 4543 aaaaaaa 2106
Qy 2099 aaaaaaa 2106

Db 4603 aaaaaaa 4610
RESULT 10
AAC78483
ID AAC78483 standard; cDNA; 4640 BP.
XX
AC AAC78483;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human PRO708 (UNQ372) nucleotide sequence SEQ ID NO:113.
XX
KW Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW expressed sequence tag; detection; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200053756-A2.
XX
PD 14-SEP-2000.
XX
PF 18-FEB-2000; 2000WO-US04341.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 12-MAR-1999; 99US-0123957.
PR 29-MAR-1999; 99US-0126773.
PR 21-APR-1999; 99US-0130232.
PR 28-APR-1999; 99US-0131445.
PR 14-MAY-1999; 99US-0134287.
PR 23-JUN-1999; 99US-0141037.
PR 26-JUL-1999; 99US-0145698.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28565.
PR 16-DEC-1999; 99WO-US30095.
PR 30-DEC-1999; 99WO-US31243.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.
XX
PA (GETH) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI Kijavini IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
XX
DR WPI: 2000-611443/58.
DR P-PSDB; AAB44257.
XX
PT Novel PRO polypeptides and polynucleotides used in detection methods,
PT to target bioactive molecules to specific cells, and to modulate
PT cellular activities -
XX
PS Claim 2; Fig 42; 636pp; English.
XX
CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed
CC sequence tag) sequences which encode secreted or transmembrane PRO
CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic
CC activity. The polynucleotides and polypeptides can be used for detecting
CC the presence of PRO polypeptides in samples, for linking bioactive
CC molecules to cells and for modulating biological activities of cells,
CC using the polypeptides for specific targeting. The polypeptide targeting
CC can be used to kill the target cells, e.g. for the treatment of cancers.
CC The polypeptide pairs provide specific targeting of bioactive molecules
CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
CC the isolation of the PRO polynucleotide sequences.
XX
SQ Sequence 4640 BP; 1427 A; 955 C; 1026 G; 1232 T; 0 other;


```
PR 07-MAR-1997; 97US-0040334.
PR 07-MAR-1997; 97US-0040336.
PR 07-MAR-1997; 97US-0040626.
PR 11-APR-1997; 97US-0043311.
PR 11-APR-1997; 97US-0043312.
PR 11-APR-1997; 97US-0043313.
PR 11-APR-1997; 97US-0043314.
PR 11-APR-1997; 97US-0043568.
PR 11-APR-1997; 97US-0043569.
PR 11-APR-1997; 97US-0043576.
PR 11-APR-1997; 97US-0043578.
PR 11-APR-1997; 97US-0043580.
PR 11-APR-1997; 97US-0043669.
PR 11-APR-1997; 97US-0043670.
PR 11-APR-1997; 97US-0043671.
PR 11-APR-1997; 97US-0043672.
PR 11-APR-1997; 97US-0043674.
PR 23-MAY-1997; 97US-0047492.
PR 23-MAY-1997; 97US-0047500.
PR 23-MAY-1997; 97US-0047501.
PR 23-MAY-1997; 97US-0047502.
PR 23-MAY-1997; 97US-0047503.
PR 23-MAY-1997; 97US-0047581.
PR 23-MAY-1997; 97US-0047582.
PR 23-MAY-1997; 97US-0047583.
PR 23-MAY-1997; 97US-0047584.
PR 23-MAY-1997; 97US-0047585.
PR 23-MAY-1997; 97US-0047586.
PR 23-MAY-1997; 97US-0047587.
PR 23-MAY-1997; 97US-0047588.
PR 23-MAY-1997; 97US-0047589.
PR 23-MAY-1997; 97US-0047590.
PR 23-MAY-1997; 97US-0047592.
PR 23-MAY-1997; 97US-0047593.
PR 23-MAY-1997; 97US-0047594.
PR 23-MAY-1997; 97US-0047595.
PR 23-MAY-1997; 97US-0047596.
PR 23-MAY-1997; 97US-0047597.
PR 23-MAY-1997; 97US-0047598.
PR 23-MAY-1997; 97US-0047599.
PR 23-MAY-1997; 97US-0047600.
PR 23-MAY-1997; 97US-0047601.
PR 23-MAY-1997; 97US-0047602.
PR 23-MAY-1997; 97US-0047612.
PR 23-MAY-1997; 97US-0047613.
PR 23-MAY-1997; 97US-0047614.
PR 23-MAY-1997; 97US-0047615.
PR 23-MAY-1997; 97US-0047617.
PR 23-MAY-1997; 97US-0047618.
PR 23-MAY-1997; 97US-0047632.
PR 23-MAY-1997; 97US-0047633.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048974.
PR 13-JUN-1997; 97US-0049610.
PR 08-JUL-1997; 97US-0051926.
PR 16-JUL-1997; 97US-0052874.
PR 18-AUG-1997; 97US-0055724.
PR 22-AUG-1997; 97US-0056630.
PR 22-AUG-1997; 97US-0056631.
PR 22-AUG-1997; 97US-0056632.
PR 22-AUG-1997; 97US-0056636.
PR 22-AUG-1997; 97US-0056637.
PR 22-AUG-1997; 97US-0056662.
PR 22-AUG-1997; 97US-0056664.
PR 22-AUG-1997; 97US-0056845.
PR 22-AUG-1997; 97US-0056862.
PR 22-AUG-1997; 97US-0056864.
PR 22-AUG-1997; 97US-0056872.
PR 22-AUG-1997; 97US-0056874.
PR 22-AUG-1997; 97US-0056875.
PR 22-AUG-1997; 97US-0056876.
PR 22-AUG-1997; 97US-0056877.
PR 22-AUG-1997; 97US-0056878.
PR 22-AUG-1997; 97US-0056879.
```

```
PR 22-AUG-1997; 97US-0056880.
PR 22-AUG-1997; 97US-0056881.
PR 22-AUG-1997; 97US-0056882.
PR 22-AUG-1997; 97US-0056884.
PR 22-AUG-1997; 97US-0056886.
PR 22-AUG-1997; 97US-0056887.
PR 22-AUG-1997; 97US-0056888.
PR 22-AUG-1997; 97US-0056889.
PR 22-AUG-1997; 97US-0056892.
PR 22-AUG-1997; 97US-0056893.
PR 22-AUG-1997; 97US-0056894.
PR 22-AUG-1997; 97US-0056903.
PR 22-AUG-1997; 97US-0056908.
PR 22-AUG-1997; 97US-0056909.
PR 22-AUG-1997; 97US-0056910.
PR 22-AUG-1997; 97US-0056911.
PR 05-SEP-1997; 97US-0057650.
PR 05-SEP-1997; 97US-0057659.
PR 05-SEP-1997; 97US-0057761.
PR 12-SEP-1997; 97US-0058785.
XX
(HUMA-) HUMAN GENOME SCI INC.
PA
Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
Feng P, Ferrle AM, Fischer CL, Florence KA, Greene JM, Hu JS;
Kyaw H, Laifleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
WPI: 1998-506364/43.
P-PSDB; AAW74744.
DR
XX
New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
disorders, immune diseases, inflammation or blood disorders
XX
Claim 1: Page 252-253; 721pp: English.
XX
This sequence represents a nucleic acid molecule designated Gene 14 from
the human cDNA clone HPMFD84 (deposited as clone ATCC 97897 and ATCC
209043) which encodes a secreted human protein. The gene can be used to
generate fusion proteins by linking to the gene to a human
immunoglobulin Fc portion (e.g. AAV59502) for increasing the stability of
the fused protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W75026)
CC which are useful for preventing, treating or ameliorating medical
conditions e.g. by protein or gene therapy. Also, pathological conditions
CC can be diagnosed by determining the amount of the new polypeptides in a
sample or by determining the presence of mutations in the new
CC polynucleotides. Specific uses are described for each of the 186
CC polynucleotides, based on which tissues they are most highly expressed in
(see AAV59511 for described uses).
XX
SQ Sequence 2323 BP; 760 A; 467 C; 438 G; 658 T; 0 other;
```

```
Query Match 4.3%; Score 91.6; DB 19; Length 2323;
Best Local Similarity 76.7%; Pred. No. 3.5e-06;
Matches 112; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
```

```
QY 1961 tgaagatttgggtattactgaacctgtacttctgtctgaatgtatgacatgatt 2020
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2176 tgattattctatgtacactgtctgtcttatttaataattatcagtgaaa 2235
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2021 catgtttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2080
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2236 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2295
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2081 aaaaaaaaaaaaaaaaaaaaaaaaaa 2106
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2296 aaaaaaaaaaaaaaaaaaaaaaaaaa 2321
```



```
XX WPI: 2000-638216/61.
DR P-PSDB; AAB00196.
XX
XX Screening drug candidates for their ability to modulate breast cancer
PT by contacting the drug to a cell expressing an expression profile gene
PT and determining modulation of expression of the gene
XX
XX Disclosure; Fig 68; 258pp; English.
XX
XX New methods for screening drug candidates are described which
CC comprise adding a drug candidate to a cell that expresses a protein
CC selected from BCL1, BCL2, BCL7, BCL1, BCL5, BCL2, BCL5, BCL2, BCL2
CC and BCL3 or their fragments and determining the effect of the drug
CC on the expression of those proteins. Antibodies to breast cancer
CC genes (specifically BCL1 or its fragment (BCL1p1 or BCL1p2)) are
CC useful for inhibiting and treating breast cancer in individuals who
CC are non-responsive to anti-oestrogen and positive for oestrogen
CC receptor. Compositions comprising BCL1 or a nucleic acid encoding
CC BCL1 are useful for eliciting an immune response in an individual.
CC The antibodies are also useful for the diagnosis and prognosis of
CC breast cancer and for screening compositions which modulate the
CC breast cancer phenotype. The method allows rapid and simple
CC detection of lymph node metastases.
XX
XX Sequence 2710 BP; 851 A; 489 C; 525 G; 845 T; 0 other;
```

Query Match 4.3%; Score 91.2; DB 21; Length 2710;
Best Local Similarity 82.0%; Pred. No. 4.1e-06;
Matches 105; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

```
QY 1978 tactgaacctgtacttctgtgaatgtatgacatgatgcattgttttaaaaaaaa 2037
   ||| ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 2583 tactacattgttaattgttaacacgtatcattataaacatttttgataaaaaaa 2642
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 2038 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2097
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 2643 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2702
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
QY 2098 aaaaaaaa 2105
   |||| ||||
Db 2703 aaaaaaaa 2710
```

Search completed: September 3, 2002, 15:55:42
Job time: 4746 sec

GenCore version 4.5		Copyright (c) 1993 - 2000 Compugen Ltd.	
OM nucleic - nucleic search, using sw model		September 3, 2002, 13:18:27 ; Search time 2627.81 Seconds (without alignments)	
Run on:		16771.108 Million cell updates/sec	
Title:		US-09-856-327-1	
Perfect score:		2106	
Sequence:		1 atcagcatgtctctctca.....aaaaaaaaaaaaaaaaaaaaa 2106	
Scoring table:		IDENTITY_NUC Gapop 10.0 , Gapext 1.0	
Searched:		1797656 seqs, 10463268293 residues	
Total number of hits satisfying chosen parameters:		3595312	
Minimum DB seq length: 0			
Maximum DB seq length: 2000000000			
Post-processing: Minimum Match 0%			
Maximum Match 100%			
Listing first 45 summaries			
Database :		GenEmbl.*	
		1: gb_ba.*	
		2: gb_htg.*	
		3: gb_in.*	
		4: gb_om.*	
		5: gb_ov.*	
		6: gb_pat.*	
		7: gb_ph.*	
		8: gb_pl.*	
		9: gb_pr.*	
		10: gb_ro.*	
		11: gb_sts.*	
		12: gb_sv.*	
		13: gb_un.*	
		14: gb_vi.*	
		15: em_ba.*	
		16: em_fun.*	
		17: em_hum.*	
		18: em_in.*	
		19: em_mu.*	
		20: em_om.*	
		21: em_or.*	
		22: em_ov.*	
		23: em_pat.*	
		24: em_pl.*	
		25: em_pi.*	
		26: em_ro.*	
		27: em_sts.*	
		28: em_un.*	
		29: em_vi.*	
		30: em_htg_hum.*	
		31: em_htg_inv.*	
		32: em_htg_other.*	
		33: em_htgo_inv.*	
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.			
SUMMARIES			
Result	Query		
No.	Score Match Length DB ID	Description	

1			
BD006657	1	BD006657	2106 bp DNA linear PAT 31-JAN-2002
LOCUS			A novel protein, a gene coding therefor and a method of using the
DEFINITION			same.
ACCESSION		BD006657	
VERSION		BD006657.1	GI:18635028
KEYWORDS		JP 03075321-T/1.	
SOURCE		Lyophyllum shimeji.	
ORGANISM		Lyophyllum shimeji.	
REFERENCE		1 (bases 1 to 2106)	
AUTHORS		Takakura, Y., Kuwata, S. and Leoue, Y.	
TITLE		A novel protein, a gene coding therefor and a method of using the	
JOURNAL		Patent: JP 03075321-T 1/16-FEB-2001;	
		JAPAN TOBACCO INC., CORPORATE JURIDICAL PERSON SOCIETY FOR TECHNO	
		INNOVATION OF AGRICULTURE YUKIO SHIMOMURA, SHIRO TAKEKAWA, NOBUO CHO	
		FORESTRY AND FISHERIES, YOSHIMITSU TAKAKURA, SHIGERU KUWATA, YASUHIRO	
		INOUE	
COMMENT		OS Lyophyllum shimeji	
		PN JP 03075321-T/1	
		PD 16-FEB-2001	
ALIGNMENTS			
BD006657 2106 bp DNA linear PAT 31-JAN-2002			
A novel protein, a gene coding therefor and a method of using the			
same.			
BD006657			
BD006657.1 GI:18635028			
JP 03075321-T/1.			
Lyophyllum shimeji.			
Lyophyllum shimeji.			
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;			
Agaricales; Tricholomataceae; Lyophyllum.			
1 (bases 1 to 2106)			
Takakura, Y., Kuwata, S. and Leoue, Y.			
A novel protein, a gene coding therefor and a method of using the			
Patent: JP 03075321-T 1/16-FEB-2001;			
JAPAN TOBACCO INC., CORPORATE JURIDICAL PERSON SOCIETY FOR TECHNO			
INNOVATION OF AGRICULTURE YUKIO SHIMOMURA, SHIRO TAKEKAWA, NOBUO CHO			
FORESTRY AND FISHERIES, YOSHIMITSU TAKAKURA, SHIGERU KUWATA, YASUHIRO			
INOUE			
OS Lyophyllum shimeji			
PN JP 03075321-T/1			
PD 16-FEB-2001			

Db 1861 TTGAGGAAGGAGCAACAGCAGTGTAAACAAACCGGTCAAGTGGCTACTTCAAGTTGAATG 1920
QY 1921 catttggtgccccaccattgtatggtgtacgatagcggttgaaagatttggattattac 1980
Db 1921 CATTCGTGGTCCCTACCATGTTGATGTGTACGATAGCGGTGGAAGATTGTTGTATTAC 1980
QY 1981 tgaacctgtactttgtctgaatagtattatggcactatgattcattgttttaaaaaa 2040
Db 1981 TGAACCTGTACTTTGCTGTAATAGTTATGGCACTATGATTCAATGTTTAAAAA 2040
QY 2041 aaaaaa 2100
Db 2041 AAAAAA 2100
QY 2101 aaaaaa 2106
Db 2101 AAAAAA 2106
RESULT 2
AR141573
LOCUS AR141573 1869 bp DNA 11linear PAT 08-AUG-2001
DEFINITION Sequence 1 from patent US 6146865.
ACCESSION AR141573
VERSION AR141573.1 GI:15101089
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1869)
AUTHORS Christensen,S., Lassen,S.Flensted and Schneider,P.
TITLE Nucleic acids encoding polypeptides having pyranose oxidase activity
JOURNAL Patent: US 6146865-A 1 14-NOV-2000;
FEATURES Location/Qualifiers
Source 1..1869 /organism="unknown"
BASE COUNT 397 a 622 c 548 g 302 t
ORIGIN
Query Match 10.1%; Score 212.4; DB 6; Length 1869;
Best Local Similarity 51.5%; Pred. No. 2.2e-33;
Matches 841; Conservative 0; Mismatches 656; Indels 135; Gaps 10;
QY 269 ggtaccacaagaagaatgaaatcgagttccagaaagatatgaccgcttcgtcaatgta 328
Db 259 GGTACACAAAGAAAGAACACCGTCGAGTACCAGAAACATCGACAAATTCGTAATGTT 318
QY 329 atcaaggagccttacacaagtctctgtctctgtcagaaaccagacgtgcctacactt 388
Db 319 ATACAAGGGCAACTTATGCCCGTCTCGGTGCCGTCAACACGATGGTTCGTGACAGCTA 378
QY 389 gatccggagcctgagcgccctcggaagtccagccatcgaacgtaaaatcct 448
Db 379 AGCCCGCGCTAGTGAAGCAAGCTTCGACG-----TCTTCGCGCAACGGGGCGAATCCA 432
QY 449 caccagcgggaattcagaaactgtctgcgagggcggttaacgcgtgagtcggcggcgatg 508
Db 433 GACCAAGACCCGTGCGCAACCTTAGTGGCCAGGCGGTACCCGCGTCTGCGCGGCATG 492
QY 509 agtaccactggagcgtctccacgcacgcgattcattcccccattgaaagtctcccgggc 568
Db 493 TCTACGACTGGACGTGCGCGACGCGCGCTTCGAGAAGCTGCAG----- 537
QY 569 atcgccgctccgaagtacgaccgcggcgagagagacgaaagagtggaaagcgtt 628
Db 538 ---CGCCCGCTGCTGTGAAGAAACGACTCAAGGCGGAGACGACCGCGGTGGGACGCTC 594
QY 629 tattccgagccgagcgtctcattcgagacttccacaaaggaattcagcaggtcaattcgg 688
Db 595 TACAAGAAGCCCGAGTCGTACTTCAAGACAGCGGCACGCCAGCCAGTTCGCCGAGTCGCCG 654

QY 689 cacaccctgttctgcgtctctttgcaagacgcgtacaaggatcgtcaacgtatctttcgc 748
Db 655 CACAACCTCGTCTCAAGAAAGCTGCAGAGGAGGTACAAAGG---GCGTGCAGCATCCAG 711
QY 749 cctctcccggttgcatgcaccggttgaagaacgcgcgggaatacgtctgaatggcactca 808
Db 712 CAGATCCCGTTCGCGGAGCGCCGAGA-----GCCCGACGTTCTGTCGAGTGGAGTCG 765
QY 809 gcaagaaatcttttccactctatctacaacgtagcaaacgagaaagagctctttaccctg 868
Db 766 GCGCACACCCGTGT-----CGATCTCGAGAACCGCCGCAACAGGACGCG 810
QY 869 ctgacgaaccactcgtgcacacgactggtccttacggcggttatgagaagaagattggc 928
Db 811 CCGAAGCAGCGCTTCAACCTCTTCCCGCGCTGCGTGCACGAACGCTGAGCGCGATAAC 870
QY 929 gctgccgaggtcaggaatctactggccaccaggaatcctagtctgcagctggacagctat 988
Db 871 GCGAATCGGAGATCGTAGGCCCTGGATGTCCGCGACCTCCACGGGGCAAGAGCATCACC 930
QY 989 atcatggcgaaggtatattgtactggcgtgggagcgatcggaacccacacagattctctat 1048
Db 931 ATCAAGGCCAAGGTGTACATCTCACCGCGCGCGGTCCACAACGCGCAGCTCTCGCG 990
QY 1049 aactcgggctt-----ctctgggtacaggttcacgcccacgaatgac--tctgtgac 1099
Db 991 GCCTCTGGATTTCGGGACGCTGGGTCTCGCGACCCCGCCAAAGCGCTCTCTGCTG 1050
QY 1100 cccaactgggaggtacatcacgagcagcgatggcattttggcagatagcttgagg 1159
Db 1051 CCTACTCTGGGACCCACATCACCGAGCAGACGCTGTCTTTCGCCAGACCGTCATGAGC 1110
QY 1160 caggaattcgtgcacagcgtgcgcacgac----- 1190
Db 1111 ACGGAGCTCATCAACAGTGTACCGGGATATGACCATTTTCGGCAAGCCGCCACCCG 1170
QY 1191 -----cttatggactgcactggtggaaa 1213
Db 1171 GACTATAGCGTCACTATACCCGGGCAACCCGAAACAACAGCACCGGACTGTGTGAAC 1230
QY 1214 gaagcgtgtgctcaacatattgccaagacccgacagatgacctgcccattcgttccgc 1273
Db 1231 GAGAAGGTGAAGAAGCACATGATGGACCCACGAGGAGACCCGCTCCCGATCCCGTTCGAG 1290
QY 1274 gatccgaaacccaggtlaacacccatttacagaagaacacccctggcacacgcagatt 1333
Db 1291 GACCTTGAGCGGAGGTACACAGCTGTTTCAGGCAACGACCCATCGGCACACCCAGATT 1350
QY 1334 caccgcatgtctttctgtacgtgcgctcgttcgttcgttcgttcgttcgttcgttcgttc 1393
Db 1351 CACCGGACGCCCTTCAGCTACGGCGCGTGCAGCAGAGCATCGACTCGCGGCTCATCGTC 1410
QY 1394 gactcgcgtggtttgacacccgacctgaagaaacacacacacacacacacacacacacac 1453
Db 1411 GACTGCGGTTCTTCGGACGACCCGACCCCAAGGAGGAGAACAGCTATGTTCTTCGGAC 1470
QY 1454 gatgttcaagcgggttacagtatgcgcgacgcgttcagatatgcacccagcactgcg 1513
Db 1471 AAGATCAGGACCGGTACAACTCCGGCAGCCGCGGTTCGACTTCGCTTCGCCGGGG-- 1528
QY 1514 tcaacgtgagagcaaggaaaaatgatggccgatgatgtgcgaagtggcgagcaacttggga 1573
Db 1529 ----GCCCGGAAGCGGAGGACATGATGACCGACATGTCGTCATGTGCGGGAAGATCGGT 1584
QY 1574 ggttatttggccacgtcccccccgagtttatggatccagcgccttgcacttcattcgtg 1633
Db 1585 GGATTCGTGCTGGGTCTTACCACAGTTTCATGGAGCCCGCTTGTCTCTGACCTTGTGT 1644
QY 1634 gggactactcgcattggcttcgacaa-----ggcaactacagtggctgaatacaactcg 1687
Db 1645 GGGACGACCGCATGGCTTCGACGAGAGAGCGGACAAAGTGTGCTGCGTACGACCGACTCA 1704
QY 1688 ctggtctgggaacttggccaaatctttatgttgcagggcaatggcaccatcaggacgggcttc 1747

Db 1705 CCGGCTTCGGCTTCAAGAACCTCTTCCTCGCGCTCGGGAACATCCCCACCGCTAC 1764
Qy 1748 ggcgagaccgacacttacgtcgatgtgcccacgctatcaagagcgcgagagcatc 1807
Db 1765 GCGGGAACCGACGCTCACCGCATGCTGCTTGGGATCAAGAGTGGAGTACATCAAG 1824
Qy 1808 aatacactcaag 1819
Db 1825 AAGAACTTCGAG 1836

RESULT 3
LOCUS I82410 1869 bp DNA linear PAT 10-JUN-1998
DEFINITION Sequence 1 from patent US 5712139.
ACCESSION I82410
VERSION I82410.1 GI:3210707
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1869)
AUTHORS Nishimura, I., Okada, K., Minamihara, T., Kawai, G., Koyama, Y. and Suzuki, M.
TITLE Pyranose oxidase, pyranose oxidase gene, novel recombinant DNA and process for producing pyranose oxidase
JOURNAL Patent: US 5712139-A 1 27-JAN-1998;
FEATURES Location/Qualifiers
source 1..1869
BASE COUNT 408 a 593 c 529 g 339 t
ORIGIN

Query Match 10.0%; Score 210.2; DB 6; Length 1869;
Best Local Similarity 50.4%; Pred. No. 6.3e-33;
Matches 857; Conservative 0; Mismatches 713; Indels 129; Gaps 9;

Qy 200 gagatcgagctgctgtagcttctacgtcttaatgccgaaggaactgcagttccc 259
Db 190 GAGCTCGTGAAGCGGGTTAAGGTGCCCATGTCGACATCGGGGAATTTGACTCTGGC 249
Qy 260 tacgtctctggtaccacaagaagaatgaatcaggttccagaagaatattgacgccttc 319
Db 250 CTGAAGATCGGTGCGCCACCAAGAAGAACACCGTCGAATACAGAGAACATTTGACAAGTTT 309
Qy 320 gtcaatgtaatacaaggagccttacaacaagtctctgttctctgtctcaagaaccagaacgtg 379
Db 310 GTGAACGTCATTCAGGGGCCAATTGATGTCTGTTTCCGTTCCTCGTCAATACCTCTCGTATC 369
Qy 380 cctacacttgatccccgggagcctggagcgcgcgccctcggaagttcagccatcgaacggt 439
Db 370 GACAGCTCAGCCGACGCTCTGGCAACCTTCATCG-----TTCCTCGTCGCAANTGGC 423
Qy 440 aaaaatcctaccagcggaattcagaaactgtctgcggagcgccgtgaacgcgtgagtc 499
Db 424 TCGAACCCAGAGCAGGACCGCTTCGTAACCTCAGTGTGTCAGGCGGTCAACGCTGCTGC 483
Qy 500 ggcgcatgagtaaccactgacgtgtctcaagccacaggaattcaccaccatggaagat 559
Db 484 GGAGCATGTCCAGCATGGACATGCGGACATGCGGACACCGCGCTTTGA----- 527
Qy 560 ctccgggcatcgccgtccgaagctcagtaacgacccggcagagagcacaagaagtg 619
Db 528 --CCGCGAGAGCGCCGCTTCTCGTGAAGGACGACGAGAGCTGACAGCCGCGAGTGG 585
Qy 620 aacgagctttattccgagcgcgagcgtctcatcgcgggacttcaccaagaattcagcag 679
Db 586 GACCGGCTGTACACCAAGCGCGAGTCACTACTTCAAGACCGGACGACCATGTTCAAGGAG 645
Qy 680 tcaattggacacaccttgctcgcgtctttgcaagacgcggtacaaagatcgtcaacgt 739

Db 646 TGGATCCGCCCAAAACCTCGTGCTCAACAAGCTCGCGAGGAATACAAAGGTACGCGCGAC 705
Qy 740 atcttgcctctcccgcttgcatgcccacgcgttgaagaacgcgcgcgaatacgtcga 799
Db 706 TTCCAGCAGATCCCGCTCGCGCAACGGTTCG-----CAGTCCGACCTTCGTCGAG 756
Qy 800 tggcactcagcagaaaaatcttttccactctatctacaacagatgacaaagcaagaagctc 859
Db 757 TGGAGCTCGGCGAACACCGTGT-----CGACCTCCAGAACAGCGCGAAC 801
Qy 860 tttaacctgctgaagaaccatcgctgcacacgactggcgcttaacggcggtatgagaag 919
Db 802 ACGGACGCGCGGAATGAGCGCTTCAACCTTCCCGCGGTTGCAATGTGAGCGCTCGTG 861
Qy 920 aagattggcgctgcgaggtcaggaatctactggccaccaggaatcctcagttcgcagctg 979
Db 862 CGCAACACGTGAACTCCGAGATCGAGAGTCTGCAATCCAGGACCTCATCTCGGCGCAC 921
Qy 980 gacagctatcatggcgaaggtatgtactggcggtggagcgtgcggaaccacacag 1039
Db 922 CGCTTCGAAATCAAAAGCAGACGTGTTCGTTCTTACAGCGCGGGCGGTCCACAACGCGCAG 981
Qy 1040 attctctataactcgggctt-----ctctgggtacaggtgcacgccacgaatgactcg 1093
Db 982 CTTCTCGTGAACCTCTGGCTTTTGGACAGCTGGGCGCGCGGACCCCGCAACCCGCGCAG 1041
Qy 1094 ttgaccccaacctggggaggtacatcacgagcagcgcgcatgttcttgcagatagtc 1153
Db 1042 TTGCTGCGCTCCCTGGGAAGCTACATCACGAGCAGTCCGCTGCTTCTGCGCAGACCGTG 1101
Qy 1154 ttgagcgaggaattcgtcgacagcgtg-----cgcgacgactccttatgactgc----- 1203
Db 1102 ATGAGCACCGAGCTCATCGACAGCGTCAAGTCCGACATGATCATCAGGGCAACCTGGC 1161
Qy 1204 -----atgg 1207
Db 1162 GATCTGGGTATACGCGTACGCTACACGCCCGCGGAGAGACCAACAAGCACCCGCGACTGG 1221
Qy 1208 tggaaagaagcgtgtctcaacatatgtccaaagaacccgcagatgcactgccattccg 1267
Db 1222 TGGAAAGAAAGGTGAAGAACCACATGATGCAGCACAGGAGGCCCTTCCCAATCCCG 1281
Qy 1268 ttccgcgctcgggaaccccggttaacaaccccatattacagaagaacacccctggcacacg 1327
Db 1282 TTCGAGGACCCCGAGCCGACGTCACCACTTGTTCAGCCATCGCACCCGTCGGCACACT 1341
Qy 1328 cagattcaaccgcgactgttttcgtaacggtgcgcgtcgtctcgtgaggtggaactcgtgctc 1387
Db 1342 CAGATTACCCGCGATGCGTTTCAGTTTACGCGCGGTGCAAGCAAGCATCGACTCAGTCTC 1401
Qy 1388 atcgtcgacctgcgctggtttggcgcaaccgacccctgaagcaaacacactttgttttc 1447
Db 1402 ATGCTGACTGGCGTTCCTCGCGGAGCGAGCCAAAGGAGGAGAAACAAGCTCTGGTTC 1461
Qy 1448 cagaacgatttcaagacgggtacagttatgcgcgagcgcagcttcagataatcgacccacg 1507
Db 1462 TCGGACAAAAATTACGGACACGTACAACATCGCGAGCGCGAGCTTCGACTTCGCGTTC 1521
Qy 1508 actg---cgtcaaacgtgagagcaagaaaatgatggccgatatagtgcgaagtggcgagc 1564
Db 1522 CGGGCCCGCAGCAGCAGGAGGCGGAGGACATGATGACCGATATGTGCGTATGTGTCGCG 1581
Qy 1565 aacttgggaggttattgcccacgctcccccccgagtttatggatccagggcttgcactt 1624
Db 1582 AAGATTGGTGGCTTCCGCGCGGTCCCTCCCGCAATTTCATGGAGCCCGGCTTGTTCCTT 1641
Qy 1625 catcttgcggggaactactcgcatttggcttcgac-----aaggcaactacagtggtgat 1678
Db 1642 CACCTCGGTGTACGACCGCATGGGCTTCGACGAGCAGGAGGACAAAGTGTGCGTCAAC 1701
Qy 1679 aacaactcgtcgtctcgtgggaactttgcaaatcttattgtgaggaatggcaccatcagg 1738
Db 1702 ACGGACGCGCGCTGTTGGCTTCAAGAACCTGTTCCTCGGTGGCTGCGGAACATTC 1761


```
Db 782 TGGAGCTCGGGGAACACCGGT-----CGACCTCCAGAACAGGCCGAAC 826
Qy 860 ttaccctgctgacgaaccatcgctgcacacgactgcttaccggtcggtatgagaag 919
Db 827 ACGGACGCGCGAATGATGACGCTTCAACCTCTTCCCGCGGTGCATGTGAGCGGTGCTG 886
Qy 920 aagattggcgtccgaggtcaagaatctactggtccacccgaatctattgctgagctg 979
Db 887 CGCAACACGCTCGAATCTCGAGATCGAGAGTCTGCACATCCACGACCTCATCTCGGGGAC 946
Qy 980 gacgctatatcatgagcaaggtatatgtactggtcgctcgagagcgtatcgcaaacacag 1039
Db 947 CGCTTCAAAATCAAGCAGAGTGTCTTACAGCCGGCGGTGCCAACAGCGCGAC 1006
Qy 1040 attctctataactcgggctt-----ctctgggtctacaggtcagccacccaatgactcg 1093
Db 1007 CTCTCTGTGAATCTGTGCTTTGGACAGCTGGCGCGCGGACCCCGGAACCCGCGCGAC 1066
-Qy 1094 ttgattcccaaacctgggaggtacatcacaggagcagccgctggtcattttccagatagtc 1153
Db 1067 TTGCTGCGCTCCCTGGGAAGCTACATCACCGAGCAGTCTGCTCTTCTGCCAGACCGTG 1126
Qy 1154 ttgaggcaggaattcgtcgacagctg-----cgagcagctcttattgactgcc----- 1203
Db 1127 ATGAGCACCAGCTCATCGACGCTCAAGTCCGACATGATCAATCAGGGGCAACCCCTGGC 1186
Qy 1204 -----atgg 1207
Db 1187 GATCTGGGTACAGCTACGTACACGCCGCGCGGAGACCAACAGCACC CGACTGG 1246
Qy 1208 tgaagaagaagctgtgtcacaatatattgcaagaaccgcagatgcactgcccattccg 1267
Db 1247 TGGAAACGAAAGGTGAAGAACCCACATGATGCAGCACCAGGAGGACCGCTTCCAATCCG 1306
Qy 1268 ttccgcatccgaaccgccaggttaaacacccatttacagaagaacacccctggcacag 1327
Db 1307 TTGAGGACCCCGAGCGCGAGGTACACCTTGTTCAGGCATCGCACCCGCTGGCACACT 1366
Qy 1328 cagattaccgcgactgttttttgcaggtgcctgctgctgaggtggactctcgtgct 1387
Db 1367 CAGATTCAACCGCATGGTTTCAGTTACGGCGCGGTGCAGCAAGCATCGACTCACGCTC 1426
Qy 1388 atcgtgcacctgcgtggttttgcgaacacgcacccctgaagaacaacacttttggtttc 1447
Db 1427 ATCGTCACTTGGGCTTCTTTCGGCGGACGAGGCCAAAGGAGGAGAAAGCTCTGGTTC 1486
Qy 1448 cagaacgatgttcaaacggtacagtatccgcagccgacttcagatatcgaaccagc 1507
Db 1487 TCGACAAAATTACGGACAGTACAACATCCCGACCGCGAGCTTCGACTTCCCGTCCCG 1546
Qy 1508 actg---cgtcaaacgtgagagaagaataatgatggccgatattgacgaagtggcgagc 1564
Db 1547 GCGGGCGCAGCAGCAAGAGCGGAGGACATGATGACCCGATATGTGCGTTATGTCGGCG 1606
Qy 1565 aacttggaggttatttggccacagtcctcccccagtttatgatcagcgccctgcaatt 1624
Db 1607 AAGATTGGTGGCTTCCTTCCCGCGCTCCCTCCGCAATTTCATGGAGCCGCTCTTGCTT 1666
Qy 1625 catcttgcgggactactcgcattgcttcgac-----aaggcaactacagtggctgat 1678
Db 1667 CACTCTGGTGTACGACCCGATGGGCTTCGACGAGCAGGAGGACAAAGTCTCGCTCAAC 1726
Qy 1679 aacaactcgtgttgggaggttggcaatctttatgttgcagggaatggccacatcagg 1738
Db 1727 ACGGACTCGCGCTGTTGGCTTCAAGAACCTGTTCTCTCGGTGGCTCGGGAACATTC 1786
Qy 1739 acgggttcggcagaaacccgacacttaoctgatgtgcacgctatcaagagcgag 1798
Db 1787 ACCGCGTACGCGGAACCCGAGCTCACGGCAATGTGCGTCTCGATCAAGAGTTGGAG 1846
Qy 1799 agcatcatcaataactca 1817
||||| ||| |||
```

```
Db 1847 TACATCAAGAAACAACCTTCA 1865

RESULT 6
LOCUS E32737 Panus cornucopeiae 1902 bp DNA linear PAT 07-FEB-2001
DEFINITION Panus cornucopeiae-derived antitumor protein and gene thereof.
ACCESSION E32737
VERSION E32737.1 GI:13017467
KEYWORDS JP 1999315096-A/1.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 1902)
AUTHORS Akihiro,M.K.I.I. and Kido,T.S.R.R.
TITLE Panus cornucopeiae-derived antitumor protein and gene thereof
JOURNAL Patent: JP 1999315096-A 1 16-NOV-1999;
NYUFUDO KURISHON GIJUTSUKENKUKUMIAI
COMMENT OS Pleurotus cornucopeiae
PN JP 1999315096-A/1
PD 16-NOV-1999
PF 07-AUG-1998 JP 1998236349
PR
PI AKIHIRO MORITA, KOJI IZUMO,KAZUAKI KIDO,TOMOHIDE SAKA, PI
RYOKO AOKI
PC C07K14/375,A61K35/84,A61K38/00,C07K16/14
CC
FH
FT
FEATURES
source Location/Qualifiers
1..1902
/db_xref="taxon:32644"
BASE COUNT 437 a 556 c 479 g 430 t
ORIGIN

Query Match 8.3%; Score 175.6; DB 6; Length 1902;
Best Local Similarity 57.1%; Pred. No. 7.5e-26;
Matches 363; Conservative 0; Mismatches 264; Indels 9; Gaps 2;

QY 1205 tggtagaagaagccgtgtgtcacaatatattgccaagaaccgcagatgcactgcccatt 1264
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1249 TGGTGAACAGAGAAGGTCCAGAAACACATGATGCAGCAGCAGGAGGATCCCTCCCGATC 1308
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1265 ccgttcgcgcatcggaaccgccaggttaaacacccatttacagaagaacacccctggcac 1324
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1309 CCCTCTGATGACCTTGAGCCACAGGTCCAGCACACTCTTCCAGGACACACATCCATGGCAC 1368
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1325 acgcagattcacgcgcatgcttttgcaggtgcctgctgctgaggtggactctcgt 1384
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1369 ACTCAGATTTCATGTCGTATGCTTTAGCTACGGCGCTGTGCGGGAGAGTATCGACAGCCGT 1428
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1385 gtcatcgtgcacctgcgtggttttgcgcacacgcagccctgaagaacaacacttttggtt 1444
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1429 CTTGTTGTTGACTGGCGCTTCTTTCGGTTCGCACCTGAACCTGTGGAGGAGAAACAAGCTGTGG 1488
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1445 ttccagaacgatgttcaagcgggtacagtatccgcagccgacttcagatatcg---a 1501
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1489 TTCTCAAAAACAGATCACTGATGATCAACCTGCCACAGCCCACTTCAGCTTCCCGTTTC 1548
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1502 cccagcactgcgtcaaacgtgagcaaggaataatgatggccgatattgtcggaagtggcg 1561
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1549 CCCCAGGGCGCACAGCCACGAGGCGGAGCTCATGATGCTGCATGTGTCACAGATGTCA 1608
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1562 agcaacttgggaggttatttggccacgctcccccccagtttatgatccagtcaccttgcga 1621
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1609 ACCAAAAGTCGTGGCTTCCTTCCCGGTTTATATCCCGAGTTTATGGCTCTCGCTGCTGTA 1668
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1622 ctctcatcttgcgggactactcgcatttgccttcgacaaggcaactacagtggc-----t 1675
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1669 CTCACCTTGGTGGAAACCCCAACCGCATGGGCTTTGATGAGGAAGCTGATAAGGCGGTGTGTC 1728
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


[illegible]


```

Db 1398 AANTGACCGGAGTCAACAATAATATTAATTTGATTATTTGGTTCTTCTCTGTAGTAGT 1457
Qy 1949 tacgatagcgctgaaagatttggtagttactgaacctgtacttggtagtgaatgattat 2008
Db 1458 TGCCAGTATATATATATATATATATATATATATATATATATATATATATATATATATAT 1517
Qy 2009 ggcactatgattcattgttttaaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 2068
Db 1518 TCATCTATTTCATATGTGAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAA 1577
Qy 2069 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaa 2106
Db 1578 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1615

RESULT 14
BC017745 1282 bp mRNA linear PRI 06-DEC-2001
LOCUS Homo sapiens, Similar to nuclear fragile X mental retardation
DEFINITION protein interacting protein 1, clone MGC:21998 IMAGE:4398678, mRNA,
complete cds.
ACCESSION BC017745
VERSION BC017745.1 GI:17389409
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Strausberg, R.
JOURNAL Direct Submission
Submitted (03-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www.shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 27 Row: d Column: 2.
FEATURES
source
1..1282
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:21998 IMAGE:4398678"
/tissue_type="Duodenum, adenocarcinoma"
/clone_lib="NIH_MGC_88"
/lab_host="DH10B"
/notes="vector: pCMV-SPORT6"
183..1013
/codon_start=1
/product="Similar to nuclear fragile X mental retardation
protein interacting protein 1"
/protein_id="AAH17745.1"
/db_xref="GI:17389410"
/translation="MHAPGMKKIKLDTPEIARWRERRRKNYPYLANIERKKLKLKLEK
EKGAVLTITTYQGMKMSHSMQAKIRSPGNKHWKNDNSRQAVTVGSGSHCLDLKL
EGPPEANADPLVLINSDSDSEKRPQHSVIPKVTPLALCSLMSSYVSGSGSESEPE
ETPTKTEADVLAEVNLDSSAPKSPQDVKATVRFSEAKSNRKKSFETNPKRKKD
YHNVTFLFEPTHPYLLLEMLAPDIRHERNVILQVCVYIIRKDFGLDTSKSKDV"

BASE COUNT 522 a 207 c 258 g 295 t

```

```

ORIGIN
Query Match 4.3%; Score 91; DB 9; Length 1282;
Best Local Similarity 78.4%; Pred. No. 1.4e-08;
Matches 109; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 1968 ttgtgttactgaacctgtacttctgtaagttagttggcactatgattcatgttt 2027
Db 1142 TATGTTGAATTTTAAAGTCCTTTCCTTGGTTCTATGCAATAAAAAATACTGATTTT 1201
Qy 2028 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 2087
Db 1202 TAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1261
Qy 2088 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 2106
Db 1262 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1280

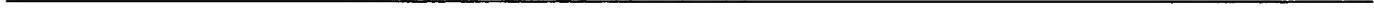
RESULT 15
AX306545 1327 bp DNA linear PAT 11-DEC-2001
LOCUS Sequence 5 from Patent WO0188139.
DEFINITION AX306545
ACCESSION AX306545
VERSION AX306545.1 GI:17645765
KEYWORDS soybean.
SOURCE Glycine max
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
REFERENCE Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
AUTHORS Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
TITLE 1 (sites)
JOURNAL Grimes, H.D., Elmer, A.M. and Murphy, K.A.
Proteins interacting with a sucrose transporter
Washington State University Research Foundation (US)
FEATURES Location/Qualifiers
source
1..1327
/organism="Glycine max"
/db_xref="taxon:3847"
208..1011
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAD19055.1"
/db_xref="GI:17645766"
/translation="MAEQLOKTFQGAPOAIPDFDNOQYFSTWQVQNQNFMTMAER
LGNALMDPMSAMLEFSNPSNRDQLEERMAKIDPDKSLHLEIETGCPAAMRY
WNNEVILQKIQAMGLNSGDASAEAGDETDIGNEDSTIVHTTASGVGVGLK
NALASGADKDEDSGRTAHFACGYEKAQVLLKAGAKVLDALDKNKTALHYAAG
YGRKECVALLLENGAAVTLQNMKGKTPIDYAKLNQNEVLKLEKDAFL"

BASE COUNT 463 a 225 c 301 g 338 t
ORIGIN
Query Match 4.3%; Score 90.8; DB 6; Length 1327;
Best Local Similarity 73.4%; Pred. No. 1.6e-08;
Matches 116; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy 1949 tacgatagcgctgaaagatttggtagttactgaacctgtacttggtagtgaatgattat 2008
Db 1157 TACCATAGATTTGTTATCTATGTTCTTGTATGATTTATTTGTACATCGATGATTTT 1216
Qy 2009 ggcactatgattcattgttttaaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 2068
Db 1217 GAAGATTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAA 1276
Qy 2069 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 2106
Db 1277 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAA 1314

```

Search completed: September 3, 2002, 15:49:51
Job time: 9084 sec



This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 3, 2002, 12:40:02 ; Search time 1745.12 Seconds
(without alignments)
16288.042 Million cell updates/sec

Title: US-09-856-327-1
Perfect score: 2106
Sequence: 1 atcagccatgtctctctca.....aaaaaaaaaaaaaaaaaaaaa 2106

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	95.4	4.5	480	9	AI392214
C 2	95.2	4.5	451	9	AL514867
C 3	94.2	4.5	165	10	BI746603
C 4	94.2	4.5	286	10	BI736597
C 5	93.8	4.5	318	10	BI941823
C 6	93.6	4.4	503	9	AL513809
C 7	93.4	4.4	162	9	AW395255
C 8	93.4	4.4	638	9	AL513901
C 9	93.2	4.4	484	10	BI680498
C 10	93	4.4	600	10	BG26622
C 11	93	4.4	685	9	AV714975
C 12	92.6	4.4	262	10	BE722980
C 13	92.6	4.4	337	10	BI142697
C 14	92.6	4.4	421	10	BG362036
C 15	92.2	4.4	308	10	BE633260
C 16	92.2	4.4	565	9	AL514205
C 17	92	4.4	175	10	BI748585

18	92	4.4	321	9	BE059229
19	92	4.4	363	10	BF612286
20	92	4.4	367	10	BI715621
21	91.8	4.4	213	9	AI496256
22	91.6	4.3	266	10	BI782468
23	91.6	4.3	464	10	BI313535
24	91.6	4.3	510	10	BE322703
C 25	91.6	4.3	926	10	BM394635
C 26	91.4	4.3	200	10	C93770
C 27	91.4	4.3	303	10	BG52798
C 28	91.4	4.3	368	9	AU176554
C 29	91.2	4.3	190	10	BM025324
C 30	91.2	4.3	383	10	BI784726
C 31	91.2	4.3	474	10	BI840173
C 32	91.2	4.3	2091	11	AF111849
C 33	91	4.3	286	10	BG737121
C 34	91	4.3	327	9	BE023867
C 35	91	4.3	342	9	AW156527
C 36	91	4.3	344	10	BI946237
C 37	91	4.3	356	10	BM367068
C 38	91	4.3	519	10	BM283661
C 39	90.8	4.3	312	9	AW317448
C 40	90.8	4.3	363	10	BM130000
C 41	90.8	4.3	603	10	BG255918
C 42	90.6	4.3	233	10	BI941734
C 43	90.6	4.3	296	9	BE057871
C 44	90.6	4.3	361	9	AU029579
C 45	90.6	4.3	456	9	AL513913

ALIGNMENTS

RESULT 1

AI392214/C

LOCUS

DEFINITION

NCSP197 Subtracted Perithecial Neurospora crassa

3' similar to nmt1 protein homolog - thiamine biosynthesis enzyme,

mRNA sequence.

AI392214

VERSION

AI392214.1 GI:4220021

KEYWORDS

EST.

SOURCE

Neurospora crassa

Neurospora crassa

ORGANISM

Neurospora crassa

REFERENCE

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

1 (bases 1 to 480)

480 bp mRNA linear EST 03-FEB-1999
NCSP197 Subtracted Perithecial Neurospora crassa cDNA clone SP197
3' similar to nmt1 protein homolog - thiamine biosynthesis enzyme,
mRNA sequence.
AI392214
AI392214.1 GI:4220021
EST.
Neurospora crassa
Neurospora crassa
Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora
1 (bases 1 to 480)
Nelson, M.A., Kang, S., Braun, E.L., Crawford, M.E., Dolan, P.L.,
Leonard, P.M., Mitchell, J., Armijo, A.M., Bean, L., Blueyes, E.,
Cushing, T., Errett, A., Fleharty, M., Gorman, M., Judson, K., Miller, R.,
Ortega, J., Pavlova, I., Perera, J., Todisco, S., Trujillo, R.,
Valentine, J., Wells, A., Werker-Washburne, M., Yazzie, S. and Natvig,
D.O.
Expressed sequences from conidia, mycelial, and sexual stages of
Neurospora crassa
Fungal Genet. Biol. 24, 348-363 (1997)
97435549
Contact: Natvig, D.O./Nelson, M.A.
Department of Biology
University of New Mexico
Cassiter Hall, Albuquerque, NM 87131, USA
Tel: 505 277 3411
Fax: 505 277 0304
Email: natvig@biology.unm.edu.
Location/Qualifiers
1. 480
/organism="Neurospora crassa"
/strain="fl a (FGSC 4347)"
/db_xref="taxon:5141"
/clone="SP197"
/clone_lib="Subtracted Perithecial"
/sex="Mating type a (fluffy), fertilized"
/tissue_type="Perithecia"

```

/dev_stage="Fruiting Body"
/lab_host="E. coli"
/Note="vector: pluescript SK (-); Site_1: EcoRI; Site_2:
XhoI; mRNA isolated from 5 day old perithecia (fruiting
bodies) of the fluffy strain fl a (Mating type a), type
fertilized with conidia from 74-OR23-IV A (Mating type
A). cDNA directionally cloned into pBluescript SK(-)
using the Uni-ZAP XR vector system (Stratagene, La Jolla,
CA). Previously identified highly expressed clones were
subtracted from this library."
BASE COUNT      128 a      137 c      48 g      166 t      1 others
ORIGIN

Query Match      4.5%; Score 95.4; DB 9; Length 480;
Best Local Similarity 67.2%; Pred. No. 0.068;
Matches 135; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 1906 actcaagtgaatcattctgtgccccctacacattgtgatgtgtacgtagcggttgaaa 1965
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 209 AGTTCAGATGAGAGATTTTGTGTGATGATGATGGGGCCCTCAGGGGTAGCTAGGAAG 150
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1966 gatttggattactgaacctgtacttctgaatgattgacacattgacattcatgt 2025
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 149 GCTTGGGTGACGTAGTAAAGTTTCTTCCAGTTTAATTGAGGTACGTTTTTAAAAA 90
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2026 ttataaaaaa 165 bp mRNA linear EST 13-FEB-2001
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 89 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 30
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2086 aaaaaa 2106
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 29 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 9

RESULT 2
AL514867/c
LOCUS      AL514867 LTI_NFL006_PL2 Homo sapiens cDNA clone CLOBB0142A12 3
DEFINITION      AL514867 LTI_NFL006_PL2 Homo sapiens cDNA clone CLOBB0142A12 3
prime, mRNA sequence.
ACCESSION      AL514867
VERSION      AL514867.1 GI:12778360
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 451)
AUTHORS      Li,W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL      Unpublished (2001)
COMMENT      Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1. .451
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CLOBB0142A12"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/Note="vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with NotI and
cloned into the NotI and EcoRV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT      200 a      76 c      16 g      132 t      27 others

```

```

ORIGIN

Query Match      4.5%; Score 95.2; DB 9; Length 451;
Best Local Similarity 65.8%; Pred. No. 0.074;
Matches 121; Conservative 10; Mismatches 53; Indels 0; Gaps 0;

QY 1923 ttctggtccccctaccattgtgatgtgtacgtagcggttgaaagattttgtgtattactg 1982
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 235 TTTTTCCTCCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 176
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1983 aacctgtacttctgaatgattgacacattgacattgacattgacattgacattgacattg 2042
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 175 TTTTTCCTCCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 116
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2043 aaaaaa 2102
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 115 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 56
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2103 aaaa 2106
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 55 AAAAA 52

RESULT 3
BI746603
LOCUS      BI746603 165 bp mRNA linear EST 25-SEP-2001
DEFINITION      xm29d01.v1 Meloidogyne arenaria egg pAMP1 v1 Chiapelli McCarter
Meloidogyne arenaria cDNA 5', mRNA sequence.
ACCESSION      BI746603
VERSION      BI746603.1 GI:15768405
KEYWORDS      EST.
SOURCE      Meloidogyne arenaria.
ORGANISM      Meloidogyne arenaria
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchida; Tylenchida;
Tylenchoidea; Heterodermidae; Meloidogyinae; Meloidogyne.
REFERENCE      1 (bases 1 to 165)
AUTHORS      McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
Dante,M., Marra,M., Hillier,L., Kucaba,T., Thelising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R.,
Ronko,I., Kennedy,S., Meguire,L., Beck,C., Underwood,K., Steptoe,
M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
Wilson,R.
TITLE      The Washington Univ. Nematode EST Project, 1999
JOURNAL      Unpublished (1999)
COMMENT      Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Brandi Chiapelli and Dr. James
McCartar (bchiapel@watson.wustl.edu & jmccarte@watson.wustl.edu) at
Washington University, St. Louis. DNA Sequencing by: Washington
University Genome Sequencing Center St. Louis. Nematodes were
provided by Dr. David Bird and Daniel Snyder of North Carolina
State University.
Seq primer: -40RP from Gibco
High quality sequence stop: 164.
FEATURES
source
1. .165
/organism="Meloidogyne arenaria"
/db_xref="taxon:6304"
/clone_lib="Meloidogyne arenaria egg pAMP1 v1 Chiapelli
McCarter"
/dev_stage="egg"
/lab_host="DH10B"
/Note="vector: pAMP1 (Gibco); Site_1: NotI; Site_2: SalI;
The library was constructed by Brandi Chiapelli and Dr.
James McCarter at Washington University, St. Louis. The
cDNA was made by using Dynabead oligo-dT priming (Dynal).

```


synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (Gibco BRL). This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelzing.

COMM EXPENDING.		
BASE COUNT	154 a	31 c 56 g 77 t

Query Match	4.58;	Score 93.8;	DB 10;	Length 318;
Best Local Similarity	61.8%;	Pred. No. 0.12;		
Matches 149;	Conservative	0;	Mismatches 92;	Indels 0;
				Gaps 0;

Qy 1866 gaaggagcaacagcagtgtaaacaaacgcgtcaagtgqctacttcaagttqaatqcattc 1925

Db 62 GAGGGTCAAGAAACAATTAAACAAAAATATCCAAGTTTCTATGACAAATTAGAAACAATT 121

Qy 1926 tggccccctaccatgttgatgtgtacgataggcgcttgaagatttttgtgtattactgaac 1985

Db 122 TGTGTACAAAAATGTTAATTTGTACATGTTGGGGTTTGGAGGTTAGCTTTATAATATAC 181

Qy 1986 ctgtactttgtctgaaataqttatqqcactatgattcatgttttaaaaaaaaaaaaaa 2045

Db 182 TTTTGTAGATAAGGTA TTGTACTTGTGTTGTTGATATAAAAAATAAAA 241

Qy 2046 aaaaaaaaaaaaaaaaaa 2105

Db

242 AAACAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 301

Ov 2106 a 2106

Db 302 A 302

RESULT	6
AL513809/c	
LOCUS	AL513809 503 bp mRNA linear EST 13-FEB-2001
DEFINITION	AL513809 LTI_NFL006_PL2 Homo sapiens cDNA clone CL0BA009ZF09 3 prime, mRNA sequence.

AL513809.1	GI:12777303
EST.	
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 503)
TITLE	Li, W. B., Gruber, C., Jessee, J. and Polayes, D.
JOURNAL	Full-length cDNA libraries and normalization
COMMENT	Unpublished (2001)
	Contact: Genoscope
	Genoscope - Centre National de Sequencage
	BP 191 91006 EVRY cedex - France
	Email: seqref@genoscope.cns.fr , Web : www.genoscope.cns.fr .

Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com>"

BASE COUNT	138 a	144 c	4 g	133 t	84 others
ORIGIN					

Query Match	4.4%	Score 93.6;	DB 9;	Length 503;
Best Local Similarity	54.8%	Fred. No. 0.1;		
Matches 92;	Conservative 40;	Mismatches 36;	Indels 0;	Gaps 0;

Qy 1939 tgttgatgtgtacgataggcgttgaaagattttgtgtattactgaacctgtactttgtct 1998

```
Dbb      215 TTTT|TTTT|TTTT|GKGG|KGKK|DKTT|TDTT|ATWAA|WWW|TTT|WWW|WWW|WWW|WT|TW  156
```

Qy 1999 qaataqtattatqgcactatqattcatqtttaaaaaaaaaaaaaaaaaa 2058

[illegible]

db 95 48

RESULT 7

AW395255	AW395255	162 bp	linear	EST 03-DEC-2001
LOCUS	sh45g11.y1	Glycine max	CDNA clone	GENOME
DEFINITION	Gm-C1017-4701.5,	mRNA sequence.		SYSTEMS CLONE ID:

VERSION
 AW395255.1 GI:6913725
 EST.
 SOURCE
 soybean.
 ORGANISM
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine
 Glycine

(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelting."

BASE COUNT 101 a 14 c 12 g 35 t
ORIGIN

Query Match 4.4%; Score 93.4; DB 9; Length 162;
Best Local Similarity 78.3%; Pred. No. 0.19; Indels 0; Gaps 0;
Matches 112; Conservative 0; Mismatches 31;

QY 1964 aagatttctgtattactgaacctgtacttctgtctgaatgattgacacatgattcat 2023
DB 20 AATAGCTTGTTGTTTACTTCTATCCCATTTTACATGTTTGGATGGAGTGGTTATTCAT 79
QY 2024 gtttaaaaaa 2083
DB 80 TTTAAAAA 139
QY 2084 aaaaaa 2106
DB 140 AAAAAA 162

RESULT 8
AL513901/c
LOCUS AL513901 LTI_NFL006_PL2 Homo sapiens cDNA clone CL0BA006ZG01 3
DEFINITION prime, mRNA sequence.
ACCESSION AL513901
VERSION AL513901.1 GI:12777395
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1. .638
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CL0BA006ZG01"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/notes="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 347 a 58 c 15 g 141 t 77 others
ORIGIN

Query Match 4.4%; Score 93.4; DB 9; Length 638;
Best Local Similarity 60.8%; Pred. No. 0.099; Indels 0; Gaps 0;
Matches 121; Conservative 16; Mismatches 62;

QY 1908 ttcaagtgtgaatgcattctgtgtccctaccattgtgtgtacgataggcgttgaaaga 1967
DB 217 TTTTWTGTAATWATTTTTTTTTTTTTTTTTTTTTTTTATTTTATTTTATTTTATTTT 158
QY 1968 ttitgtgtattactgaacctgtacttctgtctgaatgattgacacatgattcatgttt 2027
DB 157 TTTTWTWTTTTTTTACCTTTTTTTTTTTTTTTTTRAGKTTTATATAAAAWTAAATTTTTTTAT 98
QY 2028 aaaaaa 2087
DB 97 TWA 38
QY 2088 aaaaaa 2106
DB 37 AAAAAA 19

RESULT 9
BI680498
LOCUS 458776 MARC 1BOV Bos taurus cDNA 5', mRNA linear EST 17-SEP-2001
DEFINITION BI680498
ACCESSION BI680498
VERSION BI680498.1 GI:15633409
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
AUTHORS Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Perlea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keele, J.W.

TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTTCCAGTCACGACG
Plate: 127 row: L column: 3
Seq primer: ATTAGTGACACTATAG.

FEATURES
source
1. .484
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/notes="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from lymph node, ovary, fat, hypothalamus, and pituitary."

BASE COUNT 175 a 103 c 83 g 123 t
ORIGIN

Query Match 4.4%; Score 93.2; DB 10; Length 484;
Best Local Similarity 88.6%; Pred. No. 0.12; Indels 0; Gaps 0;
Matches 101; Conservative 0; Mismatches 13;

QY 1993 ttgtctgaatgattgacacatgattcatgttttaaaaaaaaaaaaaaaaaaaaaa 2052

```

Db 369 TTGCTTAACAAATTATGGAACGTGATCTATTATTTTAAACAAAAA 428
Qy 2053 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2106
Db 429 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 482

RESULT 10
BG926622
LOCUS HNC56-1-17.R HNC (Human Normal Cartilage) Homo sapiens cDNA, mRNA
DEFINITION BG926622.1 GI:14321145
ACCESSION BG926622
VERSION BG926622.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 600)
AUTHORS Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
Sathe,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
Lark,M.W.
TITLE Identification and initial characterization of 5000 expressed
sequenced tags (ESTs) each from adult human normal and
osteoarthritic cartilage cDNA libraries
JOURNAL Osteoarthritis Cartilage 9 (7), 641-653 (2001)
MEDLINE 21482651
COMMENT Contact: Sanjay Kumar
UW2109
GlaxoSmithKline
709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
Tel: 610-270-7245
Fax: 610-270-5598
Email: sanjay_kumar-legsk.com
Seq primer: T7.
FEATURES
source
1. 600
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HNC (Human Normal Cartilage)"
/tissue_type="cartilage"
/lab_host="E.coli DH10 B"
/note="Vector: pSPORT I; Site_1: SalI; Site_2: NotI;
Directional"
BASE COUNT 230 a 101 c 112 g 157 t
ORIGIN
Query Match 4.4%; Score 93; DB 10; Length 600;
Best Local Similarity 72.7%; Pred. No. 0.11;
Matches 120; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
Qy 1942 tgatgtgacatagcgctgaagattttgtgtattactgaacctgtacttgcctgaa 2001
Db 377 TGATTTATTGGAGTGGAGGCAATGTTTTTAAAGAAAAACATGTCATGTAGGTGCTAAA 436
Qy 2002 tagttatggcactatgattcatgttttaaaaaaaaaaaaaaaaaaaaaa 2061
Db 437 AATAAATGCACTTAACTCAAAAAA 541
Qy 2062 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2106
Db 497 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 496

RESULT 11
AV714975
LOCUS AV714975 DCB Homo sapiens cDNA clone DCBAYA01 5', mRNA
DEFINITION AV714975
ACCESSION AV714975
VERSION AV714975.1
KEYWORDS EST 11-OCT-2000
MEDLINE AV714975.1 GI:10796492

```

```

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 685)
AUTHORS Xu,X., Gu,J., Yang,Y., Gao,G., Xiao,H., Li,N., Qian,B., Gao,X.,
Song,H., Gu,Y., Yang,Y., Gao,G., Xiao,H., Li,N., Qian,B., Gao,X.,
Cheng,Z., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
G., Cheng,Z. and Han,Z.
TITLE Homo sapiens cDNA DCB clones
JOURNAL unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1. 685
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="DCBAYA01"
/clone_lib="DCB"
/cell_type="dendritic cells"
/dev_stage="mature"
/lab_host="BM25.8"
/note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB"
BASE COUNT 271 a 118 c 123 g 173 t
ORIGIN
Query Match 4.4%; Score 93; DB 9; Length 685;
Best Local Similarity 72.7%; Pred. No. 0.11;
Matches 120; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
Qy 1942 tgatgtgacatagcgctgaagattttgtgtattactgaacctgtacttgcctgaa 2001
Db 304 TGATTTATTGGAGTGGAGGCAATGTTTTTAAAGAAAAACATGTCATGTAGGTGCTAAA 363
Qy 2002 tagttatggcactatgattcatgttttaaaaaaaaaaaaaaaaaaaaaa 2061
Db 364 AATAAATGCACTTAACTCAAAAAA 423
Qy 2062 aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2106
Db 424 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 468

RESULT 12
BE722980
LOCUS BE722980 MARC 4BOV Bos taurus cDNA 5', mRNA
DEFINITION BE722980
ACCESSION BE722980
VERSION BE722980.1
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 262)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Pertes,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013

```


Email: est@watson.wustl.edu
Libraries were constructed by Dr. Stavros Bashiardes as part of the Physcomitrella EST program (PER) at the Univ. of Leeds (UK) and Washington Univ. in St. Louis (USA) DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Celina Knight (c.d.knight@leeds.ac.uk)
Seq primer: -40RP from Gibco
High quality sequence stop: 416.

FEATURES

SOURCE

```

    "organism": "Physcomitrella patens"  

    "db_xref": "taxon:3219"  

    "clone_name": "PEP_SOURCE ID:"  

    "clone_lib": "Woss EST library PEG"  

    "tissue_type": "ametophore: 30 day old tissue,  

    ammetaphase"  

    "accession": "D81069"

```

/base_count - 199 a 65 c 99 t

note=Vector: pAMPl; Construction of the cDNA library was performed by Dr. W. Gregg Clark using a modification of the cDNA synthesis protocol developed in the laboratory of Dr. Michael Lovett by Dr. Yulia Korshunova (personal communication). First polyA + RNA was isolated from total gametophore RNA using oligo dT magnetic beads. Following this, first strand cDNA synthesis was performed on the bead-bound polyA + RNA, during which an oligonucleotide anchor sequence was incorporated onto the 5'-ends of the cDNA. PCR amplification was then used to synthesize the second strand, to amplify the double stranded DNA, and to incorporate dUTP containing sequences into the ends of the double stranded cDNA. This DNA was size selected and cloned into pAMPl using the CloneAMP pAMPl System (Life Technologies, GibcoBRL) for cloning amplification products by a non-restriction site dependant process. The cloning was directional based on sequence asymmetry introduced at the ends during PCR amplification. The 3' cDNA ends are proximal to the NotI site of the multiple cloning site in pAMPl. This annealing mixture was transformed into chemically competent DH10B cells and selected for ampicillin resistant growth. The resulting clones (about 330,000) were pooled to make the library."

199 a 65 c 99 t

BASE COUNT

RESULT 15
BE633260
LOCUS
DEFINITION
308 bp mRNA linear EST 25-AUG-2000
uv74q09.y1 Soares mouse 3NbXS Mus musculus cDNA clone IMAGE:3412960
5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 308)
REFERENCE
AUTHORS
NCI-Cgap <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

**JOURNAL
COMMENT**

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cdapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1088772

Seq primer: -40RP from Gibco
High quality sequence stop: 293.

FEATURES

```

location/Qualifiers
1. .308
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:3412960"
/clone_lib="Soares mouse 3NDMS"
/seqs="male"

```

```

/seq_mmc
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TCTTACCAATCGATGGAGCGCGCGCGTGTGTGTGTGTGTGTGTGTGT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M. Fatima Ronaldo."

```

BASE COUNT	158 a	39 c	45 g	66 t
ORIGIN				

Query Match 4.4%; Score 92.2; DB 10; Length 308;
Best Local Similarity 70.1%; Pred. No. 0.19;
Matches 124; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

[illegible]

Search completed: September 3, 2002, 15:05:06
Job time: 8704 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 16:01:24 ; Search time 28.03 Seconds
(without alignments)
750.079 Million cell updates/sec

Title: US-09-856-327-2_COPY_76_618

Perfect score: 2901

Sequence: 1 NAEGETAVPYVPGYHKKEI.....IINTLKGGTDGKNTGEHRL 543

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	116	4.0	514	1	G6PD_MYCTU
2	101	3.5	1010	1	SCAM_RICPA
3	100	3.4	1142	1	ENAM_PIG
4	98	3.4	925	1	W70T_HUMAN
5	97.5	3.4	639	1	AMYG_ASPAK
6	97.5	3.4	1536	1	SIN3_YEAST
7	97	3.3	612	1	AMYG_ASPOR
8	96	3.3	291	1	US02_HSV11
9	96	3.3	964	1	YIN0_YEAST
10	95.5	3.3	519	1	AMYL_SACFI
11	95.5	3.3	708	1	HELS_SULSO
12	95.5	3.3	1257	1	CCAA_BACTU
13	95	3.3	500	1	YDAK_YEAST
14	94.5	3.3	773	1	CDH_PHACH
15	94.5	3.3	855	1	GAFL_SCHPO
16	93.5	3.2	406	1	YNQ5_YEAST
17	93	3.2	470	1	LEU2_AZOVI
18	93	3.2	737	1	AMY1_AEDAE
19	93	3.2	890	1	GLND_ECOLI
20	93	3.2	890	1	GLND_SALTY
21	93	3.2	955	1	VP2_BTIV17
22	93	3.2	1012	1	UBAL_SCHPO
23	92	3.2	376	1	DJBC_MOUSE
24	92	3.2	704	1	RP3A_BOVIN
25	92	3.2	813	1	KRAF_CAEEL
26	92	3.2	1584	1	U104_CAEEL
27	92	3.2	1638	1	DPO3_LACLA
28	91.5	3.2	425	1	IUCD_ECOLI
29	91	3.1	993	1	YAV1_SCHPO
30	90.5	3.1	847	1	FAV3_HUMAN
31	90	3.1	463	1	GUN_BACSP
32	90	3.1	487	1	ENGA_CHLPN
33	90	3.1	492	1	FTSA_HELPY

34	90	3.1	606	1	PRIM_MYXXA
35	90	3.1	1039	1	GUNB_CALSA
36	90	3.1	1132	1	TERT_HUMAN
37	89.5	3.1	505	1	VL1_RPVME
38	89.5	3.1	585	1	HEX3_ADR02
39	89.5	3.1	585	1	HEX3_ADE05
40	89.5	3.1	847	1	VAV3_MOUSE
41	89.5	3.1	1433	1	DPO3_BACHD
42	89	3.1	410	1	B4_USTMA
43	89	3.1	493	1	FTSA_HELPJ
44	89	3.1	498	1	RPB2_METJA
45	89	3.1	1071	1	PR16_YEAST

ALIGNMENTS

RESULT 1

G6PD_MYCTU					
ID	G6PD_MYCTU	STANDARD;	PRT;	514 AA.	
AC	O08407;				
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD).				
GN	ZWF OR ZWF2 OR RV1447C OR MT1494 OR MTCY493.07.				
OS	Mycobacterium tuberculosis.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1773;				
RP	[.]				
RN	SEQUENCE FROM N.A.				
RC	STRAIN=H37RV;				
RX	MEDLINE=98295987; PubMed=9634230;				
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,				
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala E.,				
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,				
RA	Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,				
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,				
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,				
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;				
RT	*Deciphering the biology of Mycobacterium tuberculosis from the				
RT	complete genome sequence.;				
RL	Nature 393:537-544(1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CDC 1551 / Oshkosh;				
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,				
RA	Peterson J.F., DeBoy R., Dodson R., Gwinn M.D., Haft D., Hickey E.,				
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,				
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,				
RA	Bishai W.;				
RT	*Whole genome comparison of Mycobacterium tuberculosis clinical and				
RT	laboratory strains.;				
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.				
CC	-1- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) = D-glucono-				
CC	1,5-lactone 6-phosphate + NADPH.				
CC	-1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.				
CC	-1- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE				
CC	FAMILY.				
CC	-1- CAUTION: M.TUBERCULOSIS HAS TWO GENES FOR ZWF. THIS ONE LOOKS LIKE				
CC	A CLASSICAL ZWF.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; Z95844; CAB09259.1; -				

```
DR EMBL; AE007019; AAK45757.1; -.
DR HSSP; P14111; IDBP.
DR TIGR; MT1494; -.
DR TubercuList; RV1447c; -.
DR InterPro; IPR001282; G6PD.
DR Pfam; PF00479; G6PD.1.
DR Pfam; PF02781; G6PD_C.1.
DR PRINTS; PR00079; G6PDHGRNASE.
DR ProDom; PD001129; G6PD.1.
DR PROSITE; PS00069; G6P_DEHYDROGENASE; 1.
DR Oxidoreductase; NADP; Glucose metabolism; Complete proteome.
KW ACT_SITE 210 210 BY SIMILARITY.
FT SEQUENCE 514 AA; 57343 MW; 22A9CCEDC9AB062F CRC64;

Query Match 4.0%; Score 116; DB 1; Length 514;
Best Local Similarity 22.8%; Pred. No. 0.19;
Matches 98; Conservative 48; Mismatches 163; Indels 120; Gaps 23;

QY 29 FVNVIKALQQ-VSPVVRNQNVPTLDPGAWSPAPPSSAISNGKNPHOREFENLSAEAVTR 87
Db 77 FGVVYNVAQEHCHRTFTRQONWDLAEG-FREVPGFDDDDAFQAQLAETLEKIDAE---R 132
QY 88 GYGGMSTHTWCSTPRTHPPM-RSL--PGICRPKLSNDPAEDDKWENLYSEAERLIGTST 144
Db 133 GTGGNHAFYLAIPPKSPFVVCQLHKSGLARPO-----GDRWSRVVIE----- 175
QY 145 KEFDESIRHTLVRLSLQDAKQRIFFRPLPLACHRL--KNAPEYVYVHSAENLFHSIYN 202
Db 176 KPEGHDLASARELNKAVNAVFEEAVFR-----IDHYLGKETVQNILALFANOLFDPD 231
QY 203 ----DDQKKLFTLTNRHCTRALTALTYGKKGAAE--VRN-----LLATRNPSOL 249
Db 232 AHVVDHVQ-----ITMAEDIGLGGRAGYDYGIGAARDVIQNHLMOLLALTAMEEVSFH 285
QY 250 DSYIMA-KYVVLASGAINP-----QILYNSGFSG----- 278
Db 286 PAALQAEKIKVLSATRLAEPLDQTTSRGQYAAAGWGQGVVGLLDREGFAEDSTTETFAA 345
QY 279 --LQVTPRNDSLIP-----NLGRVITEQPMAFQIVLRQEFV--DSVRDDPYGLPWWK 327
Db 346 ITLEVDVTRWAGVPPFLRGKRLGRVTEIALVF-----RRAPHLFDATMTDELG----- 396
QY 328 EAVAQHIANKPTDALIPRPDPPEQVPTTTEHPMHTQIHRDA---FSYGAV---GPE 380
Db 397 -----TNAMVIRVQ-PDEGVTLRFGSKVPGTAMEVRDVMNDFSYGSFAFADSPE 444
QY 381 VDSRVIVDL 389
Db 445 AYERLILDV 453

RESULT 2
SCA4_RICPA
ID SCA4_RICPA STANDARD; PRT; 1010 AA.
AC Q9AJ75;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Antigenic heat-stable 120 kDa protein (F5120) (120 kDa antigen)
DE (Protein PS 120) (Fragment).
OS SCA4 OR D.
GN Rickettsia parkeri.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=35792;
RN [1]
RP SEQUENCE FROM N.A.
RA Sekeyova Z., Roux V., Raoult D.;
RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
RT 'gene b' coding for an intracytoplasmic protein.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF155059; AAK30690.1; -.
KW Antigen.
FT NON_TER 1 1
FT NON_TER 1010 1010
FT SEQUENCE 1010 AA; 110694 MW; 78586D8C92FF9C5B CRC64;

Query Match 3.5%; Score 101; DB 1; Length 1010;
Best Local Similarity 18.2%; Pred. No. 7.8;
Matches 106; Conservative 88; Mismatches 243; Indels 144; Gaps 25;

QY 61 PGSSAISNGKNPHOREFENLSAEAVTRGVYGGMSTHTWCSTPR---IHPMESLPGIGRP 116
Db 55 PMSVLSGNISFSQ-----TSDPIIKAV-----RETIQPOKNDLIEQLKDLAAL--- 100
QY 117 KLSNDPAEDDKWENLYSEAERLIGT-----STKEP-DESIRHTLVRLSLQD-----AYKD 166
Db 101 -TDRLAEQKRKEIEEKEKDTLSTFFCGPANREFIDKALENPCLKKLESTIANGYN 159
QY 167 RORIFRPLPLACHRLKNAPEYVYVH---SAENLFHSIYNDKQKKLFTL----- 212
Db 160 VHTFTS---AASGYPGGPKQVQWENHVSADLRATVVKNDAGDELCTLTNETTVTKPKPT 215
QY 213 LTNHCTRALTALGGYE-----KKIGAEVRLIATR--SMVALKADGTKPSKORPVYFAH---Y 271
Db 216 LAKQDGTQVOISYSYREIDPPIKLDKADGSMHL--SMVALKADGTKPSKORPVYFAH---Y 271
QY 261 ASGAIGNPOLYNSGFSGLQVTPRNDLSLPLNGRYITEQPMAFQIVLRQEFVDSV--- 316
Db 272 EEGPNGKPKLKEISSPKPLKFAGTGDDAI---AYIEHGGEIYTLAVTRGKYKDMKVE 327
QY 317 -----RDDPYGLPWWKEAVAQHIANKPTDALIPRPDPPEQVT 354
Db 328 LNQOSVDSLQSAEDIIIGQGSKQEQLITP---QQTSSVSEPPQHKQOVPIITPNOPL 384
QY 355 TPTTEHPMHTQIH-----RDAFSYGAVG--PEVDSRVIVDLWFGAT--- 395
Db 385 QPETSQMPQSQQVNNLLNTATALSGMQDLLNLYNAGLTKEIDSKNQIDLLKEAATAIL 444
QY 396 -----DPEANLLVFQNDYQDGYSMQPQTFR-----YRPSTASNYRARRKMMAD 438
Db 445 NNEKSDIVEKQANIITAEENTVNNKNLKPDAKVAGVNAVLEIINKNDQNTPNLEKSKMLEA 504
QY 439 MCEVASNLGGYLPTSPQPM-----DPGLALHLAGTRIGFDKATTVADNNSLWDFANLY 494
Db 505 TVAVILNSENLEPKQKQMLEKAVDVGLSL-----KDDASRAATIDGKIDVVIK-SNLY 557
QY 495 VAGNGTIRTGFGENTPLTSMCHAIKSARSIINTLKGKGTGK 535
Db 558 TEDKGTMLIANGDKVNVSELSNAEQ-KLLGSVLKKGVQEQ 597

RESULT 3
ENAM_PIG
ID ENAM_PIG STANDARD; PRT; 1142 AA.
AC O97939;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Enamelin precursor.
GN ENAM.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
```



```

Db 1213 -----RLNGDLEHWF--EESLRQAYN-NKAFKLYTTIDKVTQSLVKHAHT--LMTDAK 1261
QY 228 ERKIGAAEVRNLATNPSSQLDSYIMAKVYVVLASGAIGNPQILYNSGFGSLQVTPRND 287
Db 1262 TAEIMALFVKD---RN-----ASITSKADQIYR-----LQVR-SHMS 1295
QY 288 LIPNLGR-----YITQPMACFOIVLRQEFVDSVRDDPYGLPMPWKDAVAQHTAKNPT 339
Db 1296 NTENMFRIEDKRTLHVSQYIALDDLTLEKPAEDK-----WKYYVTSYALPHPT 1347
QY 340 DALPIFRPEQVITPPTTEHPWHQIHRDAFSYGAVGPEVDSRVIVDLRFMGATDPEA 399
Db 1348 EGI-----PHEKLIKPFLE-----RLIEFGQDIDTEVDEEF-----SPEG 1383
QY 400 NNL-----LVFONDQDQ-YSM--POPTFRYRPSTASNVNRKMKMADCEVASNLG 447
Db 1384 ISVSTLIKIKIQTIVQYQHLIENSYSYVETRKATNKY-PTIANDNTQKGMVSQKELIS---- 1439
QY 448 GYLPTSPPOFMD--PGLALHLAGTTRIGFDK-----ATTVADNNSL 486
Db 1440 -----RFLDCAVGLRNLDLDEAQLSMQKKWENLKDSIAKTSAGNOCI 1481

RESULT 7
AMYG_ASPOR
ID AMYG_ASPOR STANDARD; PRT; 612 AA.
AC P36914;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Glucoamylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE (1,4-alpha-D-glucan glucohydrolase).
GN GLAA.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RX SEQUENCE FROM N.A.
RX MEDLINE=9125474; PubMed=1368680;
RA Hata Y., Kitamoto K., Gomi K., Kumagai C., Tamura G., Hara S.;
RT "The glucoamylase cDNA from Aspergillus oryzae: its cloning,
RT nucleotide sequence, and expression in Saccharomyces cerevisiae.";
RL Agric. Biol. Chem. 55:941-949(1991).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=92104497; PubMed=1761224;
RA Hata Y., Tsuchiya K., Kitamoto K., Gomi K., Kumagai C., Tamura G.,
RA Hara S.;
RT "Nucleotide sequence and expression of the glucoamylase-encoding gene
RT (glaa) from Aspergillus oryzae.";
RL Gene 108:145-150(1991).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=RIB 40;
RA Hara S., Tamura G., Kumagai C., Gomi K., Kitamoto K., Tsuchiya K.,
RA Hata Y.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -!- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D01035; BAA00841.1;

```

```

DR EMBL; D10598; BAA01540.1; -.
DR PIR; JQ1346; JQ1346.
DR HSSP; P04064; IGAI.
DR InterPro; IPR002044; CBD_4.
DR InterPro; IPR00165; Glyco_hydro_15.
DR Pfam; PF00686; CBD_4; 1.
DR Pfam; PF00723; Glyco_hydro_15; 1.
DR PRINTS; PR00736; GLHYDRLASE15.
DR PRODOM; PD001568; CBD_4; 1.
DR PROSITE; PS00820; GLUCOAMYLASE; 1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT PROPEP 20 25 BY SIMILARITY.
FT CHAIN 26 612 GLUCOAMYLASE.
FT BINDING 146 146 SUBSTRATE (BY SIMILARITY).
FT ACT_SITE 202 202 CATALYTIC BASE (BY SIMILARITY).
FT ACT_SITE 205 205 GENERAL ACID CATALYST (BY SIMILARITY).
FT DISULFID 236 239 INTERACT WITH SUBSTRATES (BY SIMILARITY).
FT DISULFID 248 475 BY SIMILARITY.
FT DISULFID 288 296 BY SIMILARITY.
FT CARBOHYD 39 39 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 612 AA; 65486 MW; CD7B23E5FA978F97 CRC64;

Query Match 3.3%; Score 97; DB 1; Length 612;
Best Local Similarity 20.5%; Pred. No. 7.6;
Matches 105; Conservative 56; Mismatches 161; Indels 190; Gaps 28;

QY 78 ENLSAEAVTRGV---GGMSTHWTCSTPRIHPMESLPGIGRPKLSNDPAEDDKENNEL 133
Db 107 EFTSSQARIQGISNPSGALSS-----GGLGEPKFNVDYTAFTGAMGRQ 150
QY 134 SEAERLIGTSTKEFDESI---RHTLVLSLQDAYKDRQIRFRLPLACHRLKNAPEYVFW 190
Db 151 RDGPALRAATAMISFGELWLVENSHTSIATDL-----VMPV-----VRNDLSYVAQ 194
QY 191 HSAENLFHSIYNDKOKKLTLLTNHRCRLALTGG--YEKKIG-----AAEVRNL 239
Db 195 YNSQSGF-DLWEEVOGTSEFTVAVSHR-----ALVEGSSFAKTVGSSCPYCDSQAPQVRCY 249
QY 240 LATRNPSSOLDYIMAKVYVVLASGAIGNPQILYNSGFGSLQVTPRNDLSLIPNLGRYITQ 299
Db 250 L-----QSFTGSGYIQAN-----FGGGRSGKDIN-----TVLGSIHTFD 283
QY 300 PMAFQOIVLRQEFVDSVRDDPYGLPMPWKDAVAQHTAKNPTDALPIPF----- 346
Db 284 POATC-----DDATFQPCARALANH--KVVTDSFRSYIATNSGRAENQAVA 328
QY 347 --RDPEPQVTTPTTEHPWH-----TOIHRDAFSYGAVGPEVDSRVIVD--LRWF--- 392
Db 329 VGRYPEDS-----YNGNPNWFLTLAAAEQLYDALYQWDKIG-----SLATDVSLPFFKAL 380
QY 393 ---GATDPEANLLVFQNDVQ-----DGYSMPQPTFRYRPSTASNVNRKMKMADCEVA 443
Db 381 YSSAATGTVASSTTVKDIVSAVKADGYQVQIVQT--YAASTGS-----MAEQ--- 427
QY 444 SNLGYLPTSPPOFMDPGLALHLAGTTRIGFDKATVADNNSLVDFANLYVAG---NGT 500
Db 428 -----YTKTDGSO-----TSARD---LTWSYAALLTANNRRNAV 458
QY 501 IRTGFGEN-----PTLTSMCHAIKARSINT 527
Db 459 VPAPWGETAATSIPIACSTTSASGYSSVVI 490

RESULT 8
ID US02_HSV11
AC P06485;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)

```

```

DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Protein US2.
GN US2.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae: Simplexvirus.
OX NCBI_TaxID=10299;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=85160822; PubMed=2984429;
RA McGeoch D.J., Dolan A., Donald S., Rixon F.J.;
RT "Sequence determination and genetic content of the short unique
RL region in the genome of herpes simplex virus type 1.";
RL J. Mol. Biol. 181:1-13(1985).
CC -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES US2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L00036; AAA96686.1; -
DR EMBL; X14112; CAA32279.1; -
DR EMBL; X02138; CAA36056.1; -
DR PIR; A05238; Q0BE72.
DR InterPro: IPR003485; US2.
DR Pfam: PF02476; US2; 1.
DR SEQUENCE 291 AA; 32470 MW; 7227D9F051FD3A78 CRC64;
SQ -----

Query Match 3.3%; Score 96; DB 1; Length 291;
Best Local Similarity 21.0%; Pred. No. 3.1;
Matches 73; Conservative 31; Mismatches 113; Indels 130; Gaps 17;

QY 236 VRNLLATRN--PSQLD-----SYIMAKYVVLASGAIGNPQILYNSGFSGLQ----- 280
DB 8 VMTLLDQNALPRTSDVSPALWSFLRLQCRILASEPLGTPVVVRPANRLRLAEPLMDLP 67
QY 281 -----VTPRNDSLIPN--LGRYTEOPMAFCQIVLRQEFVDSV-----RDPD----- 320
DB 68 KPTRPIVTRTSCRCPPNTTGLFAEDSPLE-----STEVVDVACFRLHROQPSPPRL 121
QY 321 YGLPWWKEAVA-----QHIAKNPTD-----ALPIPRDPPEQVTPP 356
DB 122 YHL--VWVGADLCVPTLEYAQRILGVRITAIKTPDAWGEPAVTRFELPENTVAWTP 179
QY 357 F--TEEHPHWHTQIHRDAFSYGAVGEVDSRVIVDLRFWFGATDPEANLLVFQNDVQDGS 414
DB 180 FPAAPNHPLETLRSRYEYQGVVLPCTNGRERDCMRWL-----RSLIALHK----- 225
QY 415 MPQTFYRSTASNVARKMADMCEVASNLGGLYFTSP---PQFMDPGALHLAHTTR 471
DB 226 -PHPA--TPGLTTTSHPVRRPCCA--CMGMPEVPDQETSPFGRGQETDPLIAVR----- 275
QY 472 IGFDKATTADNNSLVDFANLYVAGNGTIRTGTGENTPLTSMCHAI 518
DB 276 -----GERPLPHICYPV 288

RESULT 9
ID YIN0_YEAST STANDARD; PRT; 964 AA.
AC P40467;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative 108.8 kDa transcriptional regulatory protein in FKH1-STH1
DE intergenic region.
GN YIL130W.
OS Saccharomyces cerevisiae (Baker's yeast).

```

```

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrell B.G., Badcock K., Bankler A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Lewis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajadream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
CC CLUSTER DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z38059; CAA86148.1; -
DR PIR; S48404; S48404.
DR HSP; P08657; 1CLD.
DR SGD; S0001392; YIL130W.
DR InterPro: IPR001138; Zn2_Cy6_fungal.
DR Pfam: PF001172; Zn_c1us; 1.
DR PRINTS; PR00054; FUNGALZNCYS.
DR SMART; SM00066; GAL4; 1.
DR PROSITE; PS00463; ZN2_Cy6_FUNGAL_1; 1.
DR PROSITE; PS50048; ZN2_Cy6_FUNGAL_2; 1.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Nuclear protein; Zinc; Metal-binding.
FT DNA_BIND 21 47 ZN(2)-CYS(6), FUNGAL-TYPE.
FT DOMAIN 811 896 HIS-RICH.
SQ SEQUENCE 964 AA; 108780 MW; ADSABE59E4B022CC CRC64;

Query Match 3.3%; Score 96; DB 1; Length 964;
Best Local Similarity 19.5%; Pred. No. 18;
Matches 100; Conservative 76; Mismatches 163; Indels 174; Gaps 27;

QY 13 GYHKK-----NELEFQK-----DIDRFVNVKALQOQVSPVRNQNVFTLD--- 53
DB 380 GFHRKLSPNSGFSDIEIEMRKRLFTYTKLDVYNAMLGLPRISPDQDFTLP-LDLS 438
QY 54 -----PCAWSAPPGSSAISNGKNPHOREF-----ENLSAEAVTRGVGCMSPHWTCST 100
DB 439 ENITEVAYLPENQHSVLSSTGIS---NEHTKFLILNEIISLYPIKKTNSIISHTVTS 495
QY 101 --PRIHPMESLPGRPKLSN-DPAEDDKENNELYSEAEERLIGTSTKEFDESIRHTLV 157
DB 496 LELKRLNWLDSLKPLIPNAENIDPE-----YERANRLHL-----SPLHVQII 539
QY 158 RSLQDAYDKQRIFRPL-----PLACHRLKN-----APEYVE--- 189
DB 540 -----LYRPFHYLSRNMNAENVDPCLCYRRARNIAVARTVIKAKEMVSNL 587
QY 190 -----WHSAEMLPHSIY-----NDDKOKKLTLLTNHRCRLALTGYEKKIG 232
DB 588 LTGSYWTACTYTFYSVAGLLFYIHEAQLPKDSAREYDILKDAETGRSVLIQKDSMA 647
QY 233 AAERYNLLATRNPSQSDSYIMAKYVVLASGAIGNPQILYN--SGFSGLOVTPRNDSLIP 290
DB 648 ASRTYNLL-----NQIFEKLSNKTQIQT-----LHSSPSNESAFLVWNSSALKP 693
QY 291 NLGRYTEQPMACQIVLRQEF-----VDSVRDDPYGLPWWKEAVAQIAKNPTDALPIP 345
DB 694 HLGDSL-QPPVFFSSQDTKNSFLAKSEESTND-----YAMANYLNNTPISENPLN 743

```


QY 346 FRDPERQV---TTPFTEHPWHTQIHRDAFSGYGAAGVPEVDSRVIVDLRWFGATDPANNL 402
 Db 744 EAQQQDVSGQTTNSNE-----RDPNFIISDIRLDNN-----GQSNI 782
 QY 403 LVFONDV---QDGYSMPQPT---FRYRPTASN 429
 Db 783 LDATDDVFIRNDG-DIPTWNSAFDFSSKSNASN 814

RESULT 10
 AMYH_SACFI STANDARD; PRT; 519 AA.
 AC P26989; P78745;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Glucoamylase GLAI precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase) (1,4-alpha-D-glucan glucohydrolase).
 GN GLAI.
 OS Saccharomycopsis fibuligera (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycopsidaceae; Saccharomycopsis.
 OX NCBI_TaxID=4944;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KZ;
 RX MEDLINE=92137640; PubMed=1840532;
 RA Hostinova E., Balanova J., Gasperik J.;
 RT "The nucleotide sequence of the glucoamylase gene GLAI from
 Saccharomycopsis fibuligera KZ";
 RL FEMS Microbiol. Lett. 67:103-108(1991).
 RN [2]
 RP REVISIONS.
 RC STRAIN=KZ;
 RA Hostinova E.;
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 CC glucose residues successively from non-reducing ends of the chains
 CC with release of beta-D-glucose.
 CC -!- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES. ONLY
 CC DIFFERS IN 11 POSITIONS WITH GLUCOAMYLASE GLUI.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X58117; CAA41120.1; -;
 DR PIR; S14596; S14596.
 DR HSP; P08017; IAYX.
 DR InterPro; IPR000165; Glyco_hydro_15.
 DR Pfam; PF00723; Glyco_hydro_15; 1.
 DR PRINTS; PR00736; GLHYDLASE15.
 DR PROSITE; PS00820; GLUCOAMYLASE; 1.
 KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
 KW Signal.
 FT SIGNAL. 1 27
 FT CHAIN 28 519
 FT BINDING 166 166
 FT ACT_SITE 234 234
 FT ACT_SITE 237 237
 FT ACT_SITE 238 238
 FT CARBOHYD 115 115
 FT CARBOHYD 127 127
 FT CARBOHYD 205 205
 SQ SEQUENCE 519 AA; 57542 MW; A15A009A7640053C CRC64;

Query Match

3.3%; Score 95.5; DB 1; Length 519;

Best Local Similarity 19.2%; Pred. No. 7.8;
 Matches 107; Conservative 62; Mismatches 199; Indels 189; Gaps 25;

QY 20 IFKQDIDRFVNIKALQOVSPVVRNQWPTLDPCGAWAPPGSSAISGNKP----- 72
 Db 49 LDKQKDVSLYY-----LLQNIAYPEGQFNDGV--PQTVIASPSTS-----NPDYIYQWT 95
 QY 73 -----HOREFENLSAEAVTRGVGGMSTHWTCTSPRIHPHMEPLPGIGRPLK 118
 Db 96 RDSAITFLTVLSELEDNNTTLAKAVEYYINTSYNLQRTSNPSGSGFDDENHKLGEPKF 155
 QY 119 SNDPADDEKKNELYSERL-----IGTSKKEFDESIRHTLVL-----RSLQDAYKD 166
 Db 156 NTDGSAYTGAWGRPQNDGAPALRAYAISRYLNDVNSLNKGLVLTDSGDINFSTEDIYKN 215
 QY 167 RQIRFPLPLACHRLKNAPEYV--EWSAENLFHSTYNDKOKKLTLLTNHRCRTPLALT 224
 Db 216 ---ILKP-----DLEYVIGYVDSYG---FDLWEENQGRHFTSLVQOKA---LAYA 257
 QY 225 GGYEKKIGAAEVRNLLATRNPSQSDSYIMAKVYVILASGAIGNPQILYNSGSGLOVTPR 284
 Db 258 VDIAKSFDDGDFANTLS--STASTLESYL-----SGSDG-----GFVNTDV--- 296
 QY 285 NDSLTPNLGRYITEQPMAPCQIVLRQEFVDSVRDDPYGLPWWKEAVAQIAHNPTDALPI 344
 Db 297 -----NHIVENPDLLQ--NSRQGLDS-----ATYIG-----PL 323
 QY 345 PFRDPEPQVTTPTTEHPWHTQIHRDAFSGYGAAGVPEVDSRVIVDLRWFGATDPANNLV 404
 Db 324 LTHDIGESSSTPF-----DVDNEYVL-----QSYLLLL 351
 QY 405 FQN-----DVQDGYSPQPTFRYRPTASNVRARKMADMCEVASNLGGYLPTSPQFMDP 460
 Db 352 EDNKORYSVNSAYSAGAAIGRY-PEDVYN-----GDSSSEGNPWFAT 393
 QY 461 GLALHLAGTTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRTGEGENTPLSMCHAIS 520
 Db 394 AYAAQVPYKLVVDYDAKSASNDITINKYDFNFKYIVDLSTINSY-----QS 440
 QY 521 ARSIINTLKGSDGKNT 537
 Db 441 SDSV--TIKSGSDEFNT 455

RESULT 11
 HELS_SULSO STANDARD; PRT; 708 AA.
 ID HELS_SULSO
 AC Q97YI9;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Putative skl2-type helicase (EC 3.6.1.-).
 DE SS02462.
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 OX NCBI_TaxID=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=21332296; PubMed=11427726;
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Aweyer M.J., Chan-Welher C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.B., Doolittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Sengen C.W., Van der Oost J.;
 RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 CC -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. SKI2 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE006844; AK42601.1; ALT_INIT.
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR003583; HHH_1.
 CC InterPro: IPR001650; Helicase_C.
 CC Pfam: PF00270; DEAD; 1.
 CC Pfam: PF00271; Helicase_C; 1.
 CC SMART: SM00487; DEXDC; 1.
 CC SMART: SM00490; Helic; 1.
 CC SMART: SM00278; HHH1; 1.
 CC Hypothetical protein; Hydrolase; Helicase; ATP-binding;
 CC Complete proteome.
 CC NP_BIND 46 53 ATP (POTENTIAL).
 CC SITE 145 148 DEXH BOX.
 CC SEQUENCE 708 AA; 80500 MW; D52C24C89541C780 CRC64;

Query Match 3.3%; Score 95.5; DB 1; Length 708;
 Best Local Similarity 21.3%; Pred. No. 12;
 Matches 53; Conservative 40; Mismatches 93; Indels 63; Gaps 11;
 QY 69 GKPHOREFENLSAEAVTRGVGGMSTHTCTSPRIH-----PPMESLPGIGRKLNSNDPAE 124
 Db 492 GKRVDLYINPFTADIIRKLEGYKA--SCEIAYLHLATPDGPIVSGR-----N 541
 QY 125 DKWENLYSEAE-RLTGTSTKEFEDEIRHTLVLRSLQDAY-----KDRQIRFLPL 175
 Db 542 EEEELIELLEDCELVVERPYEDE--YSLYNALKVALIMKDWIDEVDEDTILGKYN 598
 QY 176 LACHRLKNAPEYEW--HSAENLFHSIYNDKQKFLTLNHRCTRLALTGGYKKIGA 233
 Db 599 IGSGLRNIYETMDWLYTSAYHLSKELRDLDDHDKL-----RLNLRVTDGVEEL-- 649
 QY 234 AEVRNLLATRNPSQLDSYIMAKVYVILASGAIG--NPQILYNGSGFGLQVTPRNDSLIPN 291
 Db 650 -----LELVQIGVGGRKRRLLYNNGIKGLGDVWVNPDRVN 686
 QY 292 -LGRYTEQ 299
 Db 687 LLGQKLGER 695

RESULT 12
 ID CCAA_BACTU STANDARD; PRT; 1257 AA.
 AC Q45754;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Pesticidal crystal protein cry12Aa (Insecticidal delta-endotoxin
 DE CryXIIa(a)) (Crystalline entomotoxin) (142 kDa crystal
 DE protein).
 GN Cry12AA OR CRYXIIA(A) OR CRYVB.
 OS Bacillus thuringiensis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1428;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 2-11.
 RC STRAIN=NRRL B-18244 / PS33F2;
 RA Narva K.E., Payne J.M., Schwab G.E., Hickie L.A., Galasan T.,
 RA Nick A.J.;
 RT "Novel Bacillus thuringiensis microbes active against nematodes, and
 RT genes encoding novel nematode-active toxins cloned from Bacillus
 RT thuringi.";
 RL Patent number EP0462721, 27-DEC-1991.
 CC -1- FUNCTION: ENDOTOXIN WITH NEMATOCIDAL ACTIVITY.
 CC -1- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING

CC SPORULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART
 CC OF THE SPORE COAT.
 CC -1- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
 CC N-TERMINUS.
 CC -1- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L07027; AAA22355.1;
 CC InterPro: IPR001178; Endotoxin.
 CC Pfam: PF00555; endotoxin; 2.
 CC Toxin; Sporulation.
 CC SEQUENCE 1257 AA; 142265 MW; 3D988FFC6C0E3981 CRC64;
 Query Match 3.3%; Score 95.5; DB 1; Length 1257;
 Best Local Similarity 20.7%; Pred. No. 29;
 Matches 120; Conservative 78; Mismatches 206; Indels 175; Gaps 34;
 QY 12 PGYHKNEIEFKDIDRFVNVIKALQOVSVVRNQNVPT-LDPGAWSPGSSAISNGK 70
 Db 359 PNTHETVHVFQSYNTQSSGNISRGSNPIPIDLNPIISTCIRNSFYKATGSSVLNFK 418
 QY 71 NPHOREFENLSAEAVTRGVGGMSTHTCTSPRIHPPHESLPGIGRKLNSNDPAEDDEKN 130
 Db 419 DGTQ---GYAFQAQPTGGA-----WD-----HSFIES-----DGAPEGHKLN 452
 QY 131 ELISEAERLTGTSKEFDESIRHTLV-LRSLQDAYKQRIKRLPLACHRLKN----- 183
 Db 453 YITSP---GDLRDFINV--YTLISTPTINELSTKIKGF---PAEKGYIKNOGIMKY 503
 QY 184 --APEYVHSAENLFHSIYNDKQKFLTLNHRCTRLALTGGYKKIGAAEVRNL-- 239
 Db 504 YGKPEYINGAQPVNL-----ENQOTLIFEF-----HASKTAQYIRIRYASTQGTGK 550
 QY 240 -----LATRNPSQLDSYIMAKVYVILASGAIGNPQILYNGSGFGLQVTPRNDSLIPN 291
 Db 551 YFRDNOELQTLNPTSHNGYV-----TGNIGENYDLYTIG--SYTITEGNHTL--- 597
 QY 292 LGRYTEQPMACQI-----VL-ROEFV--DSVRDDPYGLPWKEAQAHIKAKNPTD 340
 Db 598 -----QIHNDRKNGVLDRIEFVPRKSLQDSP----- 624
 QY 341 ALPIPRDPEPQV---TTPFTEHP--W-----HTQIHRDAFSYGAVPEVDSRVIVDL 389
 Db 625 -----QDSPPEVHESTIIFDKSSPTIWSNKHSHYSHIHLEG-SYTSQG-SYPHNLINL 676
 QY 390 RWFQATPEANLLVFON-DVQDGYSMQPTFRYPSTASVNRARKMADMCCEVASNLG- 447
 Db 677 --FHPTDNRNHTIHVNGDMNDYD-----KDSVADGLNFKNTATIPSDAWSGT 726
 QY 448 -----GYLTPSPQFMDPGDLALHLAG--TTRIGFDKAT---TVADNNSLVDFANL 493
 Db 727 ITSMHLENDNNFKITPKF---ELSNELNITTOVNALFASSAQDTLASVSDYTW-IEQV 782
 QY 494 YVAGNGTIRTFGF-ENPTLTSMCHAISRSIINTLKGK 531
 Db 783 VMKVDALSDVEFGKEKKALKRLVNAQKRLSKIRNLLIGG 821
 RESULT 13
 ID YDAK_YEAST STANDARD; PRT; 500 AA.
 AC P28817;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)

Search completed: September . 3, 2002, 16:01:29
Job time: 338 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 15:58:45 ; Search time 54.14 Seconds
(without alignments)
963.733 Million cell updates/sec

Title: US-09-856-327-2_COPY_76_618
Perfect score: 2901
Sequence: 1 NAEETAVPYVGYHKKEI.....IINTLKGGTGGKNTGEHRML 543

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	170.5	5.9	573	2 B81385	probable oxidoredu
2	166.5	5.7	579	2 B87451	oxidoreductase, GM
3	159	5.5	615	2 B38575	gluconate 2-dehydr
4	158.5	5.5	529	2 A84260	hypothetical prote
5	158.5	5.5	619	2 AB2003	dehydrogenase chal
6	144	5.0	722	2 B75608	GMC oxidoreductase
7	141	4.9	591	2 H83362	gluconate dehydrog
8	140.5	4.8	562	2 JC7628	glucoside 3-dehydr
9	135	4.7	561	2 E98192	probable oxidoredu
10	135	4.7	561	2 AE3094	oxidoreductase Atu
11	128.5	4.4	748	2 T10651	hypothetical prote
12	122	4.2	527	2 E98202	dehydrogenase chal
13	122	4.2	527	2 AD3084	dehydrogenase Atu4
14	116	4.0	502	2 T18562	hypothetical prote
15	116	4.0	514	2 B70917	probable zwf2 prot
16	116	4.0	599	2 T19711	hypothetical prote
17	114	3.9	1012	2 T00958	hypothetical prote
18	113.5	3.9	578	2 F70736	probable chob prot
19	111.5	3.8	494	2 AD3255	1-sorbose dehydrog
20	110.5	3.8	589	2 T50698	probable mandeloni
21	110	3.8	1498	2 AF1082	B. subtilis Yuka p
22	108	3.7	1125	2 T19193	hypothetical prote
23	106.5	3.7	775	2 B72074	hypothetical prote
24	106.5	3.7	775	2 C81594	hypothetical prote
25	106.5	3.7	775	2 D86549	hypothetical prote
26	105.5	3.6	1611	2 T38236	hypothetical prote
27	105	3.6	499	2 T45749	hypothetical prote
28	104	3.6	1917	2 C88728	protein C48A7.1 [I
29	102.5	3.5	768	2 JC6364	cellobiose oxidase

30 102 3.5 1018 2 T40253
31 102 3.5 1042 2 T16169
32 102 3.5 2361 2 T25752
33 101.5 3.5 489 2 C82241
34 101.5 3.5 539 2 C81805
35 101.5 3.5 747 2 F87263
36 100.5 3.5 517 2 G81080
37 100 3.4 1142 2 T37455
38 99.5 3.4 501 2 B81048
39 98.5 3.4 770 2 S60676
40 98 3.4 625 2 T20634
41 98 3.4 639 2 T43190
42 98 3.4 7576 2 T17428
43 97.5 3.4 328 2 AF0170
44 97.5 3.4 569 2 E86957
45 97.5 3.4 578 2 D95400

ALIGNMENTS

RESULT 1
E81385
Probable oxidoreductase chain Cj0415 [imported] - Campylobacter jejuni (strain NCTC 1
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C:Accession: E81385
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: A81250; MUID:20150912
A:Accession: E81385
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-573 <PAR>
A:Cross-references: GB:AL139075; GB:AL111168; NID:96967817; PIDN:CAB74251.1; PID:9696
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj0415

Query Match 5.9%; Score 170.5; DB 2; Length 573;
Best Local Similarity 20.9%; Pred. No. 4.9e-05;
Matches 111; Conservative 61; Mismatches 237; Indels 123; Gaps 21;
QY 77 FENLSAEAVTRGVGSMSTHTWCSTPRIHPMESLPGICRP----KLSNDPAEDD-----K 127
DB 84 YRKMGSFLLGNVGGAGVHWNGWTFRFPYDFEIQTLKQRYGNKLGNDYTLQDMGVITYK 143
QY 128 EWNELYSAEERLIGTSTKEF-----DESRHTVLVRSLODAYKDRQRIFR 172
DB 144 DMEPYDRFEKTCGVSGEPNPLAEKMGAFRSPYPOEPLNTKMLKRFESAASNNLHY 203
QY 173 PLPLA-----CHRLKNAPEYVEHSAENLFHSIYNDKQKKLF 210
DB 204 RLPASNSKGGYTNPDGDLAPCOYCAYCERFG-----CEYGAKASPLNTVPKAMSTCKY 258
QY 211 TLLNHRCTRLLALGGYKKIGAAEVRNLLATRNPSOLDSDYIM-AKVYVLASGAIGNPQ 269
DB 259 TIRYSNVNTOIL-----KDKGVTGVKFDVTRT-----MKEYIQPADIVLTYSYMFNNAK 308
QY 270 ILYNSGSGGLQVTP-----RNDSLIPNLG--RYITPEQPMACQIVLRQEFVDSVRDD 319
DB 309 LLAYSNI-GEQYDPTKGTGKTLGRNYCYOMNMGTTAFFDEQNTFF-----MGSGALGTTSD 363
QY 320 PYG--LPWWKE-----AVAQHIAKNPTDALPIFRDPEQVPTTPTTEHPWHITQIHRD 370
DB 364 FNGDNFDSKSEKFLHGAMIYSVQLGTRPIQSAPLPAGAP-----TWGAEEF-KK 410
QY 371 AFSVG-----AVGPEVDSRVIVDLRFWFCATDDEANLLVFQNDVQDGYSMQPTFRYRPS 425
DB 411 ALNYNFTRAITVGGQ-----GASLPHKNNYLSLDPTYKDAFGMPLLRLTYN-F 457

hypothetical prote
hypothetical prote
hypothetical prote
GGDEF family prote
TspB protein NMA17
TspB domain protein
TspB protein, prob
enamelin precursor
cellulose oxidase
TspB protein, prob
hypothetical prote
probable actin-bin
FK506 polyketide s
tetraacyldiacchar
probable cholesterol
probable oxidoredu

C;Accession: B75608
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
A;Reference number: A75250; MUID:20036896
A;Accession: B75608
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-722 <WHI>
A;Cross-references: GB:AE001862; GB:AE001868; NID:g6460468; PIDN:AAF12230.1; PID:g646052
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRA0127
A;Map position: 2

[illegible]

RESULT 7
H83362
gluconate dehydrogenase PA2265 [imported] - *Pseudomonas aeruginosa* (strain PAO1)
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83362
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic patho
A:Reference number: AB2950; MUID:20437337
A:Accession: H83362
A>Status: preliminary

A:Molecule type: DNA
A:Residues: 1-591 <STO>
A:Cross-references: GB:AE004652; GB:AE004091; NID:g9948287; PIDN:AAG05653.
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2265

Query Match 4.9%; Score 141; DB 2; Length 591;
Best Local Similarity 19.9%; Pred. No. 0.0096;
Matches 120; Conservative 63; Mismatches 235; Indels 184; Gaps

Qy 45 RNONVPTDPCGAW-----SAPGGSSAISGNKPNHQREPNLSAEAVTR 87
||| |||
Db 38 RGENRDYTPDGAYPNTLDLTNTYTRGKLFQNLKSSTVSIIRGHINDYALPYRQLSALPGD 97
||| |||
Qy 88 GVGGMSTHTWCSTPRRIHPMESLPG-----IGRPKLSNDPAEDDKENWELYSEAE 137
||| |||
Db 98 GVGAGLHWSGVHFRIMPEELRLRSHYEERYGKKFIPEGMTIDQYGVSYSELEPHDFAE 157
||| |||
Qy 138 RLIGTSKEFDESIRHTL-----VLRSLDADYKDRQRIER----- 172
||| |||
Db 158 KVFQTS-----GTAHTYKGVGVKGPNFFADRSDDPPLPALQOVYS--AQLFRKAAEEL 209
||| |||
Qy 173 -----PLPLA-----CHRLKNAPEYVEHSAENLPHSIYNDKQKK 208
||| |||
Db 210 GLHPYDLPAANASGPNWNPYGVQGPCNFCGFCGSGYACYWYKSNPNL--NILPALRQTP 267
||| |||
Qy 209 LFTLLTHRCRTRALATGGYEKKIGA----AEVRNLLATRNPPSSQLDSYIMAKYVYVLASGA 264
||| |||
Db 268 LFE LRANCVLKVNLSDGQATGVTYVDAQGREIVQP-----AKLVIISAFQ 315
||| |||
Qy 265 IGNPQILYNSGFSGLQVTPRNDLSLIPNLGRYITEOPNAFCQIVLROEFVDSVRDDPY-- 321
||| |||
Db 316 FHNVRLLLSGR-GKPYDPRTEGV--VGKNFAYQNNM-----TIKAFFDKRDVHTNPFVGT 368
||| |||
Qy 322 -----GLPWKEAQAQIAKKNPTDALPIPRDRPEPQVTTPE 357
||| |||
Db 369 GGGGVAVDDFNADNFDHGPLGFGVGGSPW---VQAGSRPIGGLAVPQGTPS----- 417
||| |||
Qy 358 TEHPWHT---QIHRDAFSYGAVGPEYDSRVIVDLRWFAGTDPPEANLLYFQNDVDGY 414
||| |||
Db 418 ---WGSQWQAVKDAYTH-----TVSMDAHGSNMTYRNDYLDLDPTYKDAYG 461
||| |||
Qy 415 MP--OPTERY-----RPTASNVPRKMMADMCEVA-----SNLGGYLPSPPOFMDPGL 462
||| |||
Db 462 QPLLRTMFDKMDKNEIRSRVYVTEHMRKIAEAMNPKALSVSVKNFGDHFNTRVYQ----- 515
||| |||
Qy 463 ALHLAGTTRIGDKATTVADNNSLVDFANLYVAGNCTIRTGEGENPTLITSMCHAIKSAR 522
||| |||
Db 516 TTHLLGGAINGSDPKTSVLNRYLQSDVDHNVFYMGASAFPGQGTGYNPTGLVLAALATWSAK 575
||| |||
Qy 523 SI 524
||| |||
Db 576 AI 577

RESULT 8
JC7628
glucoside 3-dehydrogenase (EC 1.1.99.13) - Halomonas sp. alpha-15
C:Species: Halomonas sp. alpha-15
C:Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001
C:Accession: JC7628; PC7123
F:Kojima, K.; Tsugawa, W.; Sode, K.
Biochem. Biophys. Res. Commun. 282, 21-27, 2001
A:Title: Cloning and expression of glucose 3-dehydrogenase from Halomonas
A:Reference number: JC7628; MUID:21164693; PMID:11263965
A:Accession: JC7628
A:Molecule type: DNA
A:Residues: 1-562 <KOJ>
A:Accession: PC7123
A:Molecule type: protein
A:Residues: 2-11 <KO2>

Db 450 KRCVGVFASVGGKIGK-----FIIEARVTVSAGSLTPTPLMSGLKN-----495
QY 286 DSLPLNLR-----YTEOPMAFCQIVLRQEFVDSVR--DDPYGLPWNKEAVA 331
Db 496 ----PNIGRNKLHPVLTWGFPEKDSFGKNYEGGIITSVHHND-----TESGC 544
QY 332 QHTAKNPTDALPTPFRRDPEPQVTPPTTEBHPWHTQIHRDAFSYGAGPEVDSRVIVDLRW 391
Db 545 KATLENPLIG-PASVAGLSPWVS-----GPDKERMIR-----576
QY 392 FGATDPEANNLLVFQNDVODGYSM--PQTFYRSTASNVRARKMADMCVAS----NL 446
Db 577 YGWT-----AHLFALVRDLGSGEVMENEVYTRTKDRENLRAGLRQALRVSAAGAVEV 632
QY 447 GGV-----LPTSPPOFMDPGLAL-----HLAGTTRIGFDKAT 478
Db 633 GYRSDGQRMKCEAITKEAMEFELDEVDAGVGVGKGEYWTYFSAHQMSRWGVTAE 692
QY 479 TVADNNSLVMDFANLYVAGNGTIRGTGNGENPTLTSMCHAISARSINTLKGT 532
Db 693 GALLDENGESWEAEGFLVCDGSLPSAVGVNPMITIQSTAYCISKIVDSIQNK 746
RESULT 12
dehydrogenase chain I (AF068066) [imported] - Agrobacterium tumefaciens (strain C58, Cera
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: E98202
R:Goodner, B.; Hinkie, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: E98202
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-527 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89143.1; PID:g15158953; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L1138
A:Map position: linear chromosome

Query Match 4.2%; Score 122; DB 2; Length 527;
Best Local Similarity 19.8%; Pred. No. 0.23;
Matches 101; Conservative 55; Mismatches 175; Indels 180; Gaps 22;
QY 75 REFENLSABAVTRGVGMSTHWCSTPRTHP---PMESLPGIGRPKLSNDPAEDDKWNE 131
Db 104 RDRFEVEHE-----DGVSPAMPVSYAELEPFYAEARLFGV-RGRAGDDPTPEPRSA 156
QY 132 LYSEA--ERLIGTSKEPDESIRHTLVLSLQDAYKDRQIRFRPLA-----CHRL 181
Db 157 MHAPIHEPVGKAGFER-----LGLRPFH-----MPSAIDYGPGLCRR 199
QY 182 KNAPEYVWHSANLFSHYNDKOKKLF-----TLNHRCTRALTGGYEKKIG 232
Db 200 GTCDAFVCFDAKG-----DAETRLRLPALRHPNVSLLTGARVRL--IADGGKHIV 250
QY 233 AAEVNRLLATRNPPSOLDOSYIMAKYVVLASGAIGNPQILYNSGSLQVTPRNDSLIPN- 291
Db 251 AVEIER-----AGEITT-IEAPLFLVSAGAINSALLRSA-----DEKPKNG 292
QY 292 -----LGRYTEQPM-----FCQIVLRQEFVDSVRDPP-----320
Db 293 LANSVGVGRYLMNHLGSLGMLLPFTINDTRFPKTMNSLNDFFDGTGDEAARGNVQMLG 352
QY 321 -----YGLFWKAEVAQHIKAKNPTDALPIFRDPEPQVTPPTTEHPWHTQIHRDAF 372
Db 353 NIQGPWIRAAVPMWRPLANLLARHSVDFLVN-----SED-----387
QY 373 SYGAVGPEVDSRVIVDLRWFGATDPEANNLLVFQNDVODGYSMPQTFYRSTASN--- 429
Db 388 ----TPKYDSRV-----KPGWKNAG-----LIYRPGDREAHOR 417
QY 430 -VRARKMADMCVASNLGGYLTSPPOFMDPGLALHLAGTTRIGFDKATTVADNNSLVW 488
Db 418 FVRHMRSLRLKNGFPVVLGHSGFIEAPS-----HOCGTVMRGDDPKAEALNALCOTY 469

QY 373 SYGAVGPEVDSRVIVDLRWFGATDPEANNLLVFQNDVODGYSMPQTFYRSTASN--- 429
Db 388 ----TPKYDSRV-----KPGWKNAG-----LIYRPGDREAHOR 417
QY 430 -VRARKMADMCVASNLGGYLTSPPOFMDPGLALHLAGTTRIGFDKATTVADNNSLVW 488
Db 418 FVRHMRSLRLKNGFPVVLGHSGFIEAPS-----HOCGTVMRGDDPKAEALNALCOTY 469
QY 489 DFANLYVAGNGTIRGTGNGENPTLTSMCHAIS 519
Db 470 DHPNLYVDAGFFPSAALNPALTVAQAIR 500
RESULT 13
AD3084
dehydrogenase Atu4296 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AD3084
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AD3084
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-527 <KUR>
A:Cross-references: GB:AE008689; PIDN:AAL45090.1; PID:g17742757; GSPDB:GN00187
C:Genetics:
A:Gene: Atu4296
A:Map position: linear chromosome

Query Match 4.2%; Score 122; DB 2; Length 527;
Best Local Similarity 19.8%; Pred. No. 0.23;
Matches 101; Conservative 55; Mismatches 175; Indels 180; Gaps 22;
QY 75 REFENLSABAVTRGVGMSTHWCSTPRTHP---PMESLPGIGRPKLSNDPAEDDKWNE 131
Db 104 RDRFEVEHE-----DGVSPAMPVSYAELEPFYAEARLFGV-RGRAGDDPTPEPRSA 156
QY 132 LYSEA--ERLIGTSKEPDESIRHTLVLSLQDAYKDRQIRFRPLA-----CHRL 181
Db 157 MHAPIHEPVGKAGFER-----LGLRPFH-----MPSAIDYGPGLCRR 199
QY 182 KNAPEYVWHSANLFSHYNDKOKKLF-----TLNHRCTRALTGGYEKKIG 232
Db 200 GTCDAFVCFDAKG-----DAETRLRLPALRHPNVSLLTGARVRL--IADGGKHIV 250
QY 233 AAEVNRLLATRNPPSOLDOSYIMAKYVVLASGAIGNPQILYNSGSLQVTPRNDSLIPN- 291
Db 251 AVEIER-----AGEITT-IEAPLFLVSAGAINSALLRSA-----DEKPKNG 292
QY 292 -----LGRYTEQPM-----FCQIVLRQEFVDSVRDPP-----320
Db 293 LANSVGVGRYLMNHLGSLGMLLPFTINDTRFPKTMNSLNDFFDGTGDEAARGNVQMLG 352
QY 321 -----YGLFWKAEVAQHIKAKNPTDALPIFRDPEPQVTPPTTEHPWHTQIHRDAF 372
Db 353 NIQGPWIRAAVPMWRPLANLLARHSVDFLVN-----SED-----387
QY 373 SYGAVGPEVDSRVIVDLRWFGATDPEANNLLVFQNDVODGYSMPQTFYRSTASN--- 429
Db 388 ----TPKYDSRV-----KPGWKNAG-----LIYRPGDREAHOR 417
QY 430 -VRARKMADMCVASNLGGYLTSPPOFMDPGLALHLAGTTRIGFDKATTVADNNSLVW 488
Db 418 FVRHMRSLRLKNGFPVVLGHSGFIEAPS-----HOCGTVMRGDDPKAEALNALCOTY 469

Qy 489 DFANLYVAGNCTIRTFGENFTLSTMCCHAIK 519
| | | | | : : : : :
Db 470 DHPNLYVVDAGFFPSSAALNPALTVAAQALR 500
| | | | | : : : : :

RESULT 14

T18562

hypothetical protein F15C11.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999

C:Accession: T18562; T20970

R:Barlow, K.

submitted to the EMBL Data Library, August 1997

A:Reference number: Z18976

A:Accession: T18562

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-502 <W1>

A:Cross-references: EMBL:Z98262; PIDN:CAB10932.1; GSPDB:GN00019; CESP:F15C11.2

A:Experimental source: clone VF15C11

R:Wilkinson, J.

submitted to the EMBL Data Library, April 1996

A:Reference number: Z19352

A:Accession: T20970

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-502 <W12>

A:Cross-references: EMBL:Z71260; PIDN:CAA95799.1; GSPDB:GN00019; CESP:F15C11.2

A:Experimental source: clone F15C11

C:Genetics:

A:Gene: CESP:F15C11.2

A:Map position: 1

A:Introns: 31/3; 67/1; 169/3; 212/3; 293/2; 294/1; 340/3; 368/3; 406/2; 428/2; 448/3; 48

Query Match 4.08; Score 116; DB 2; Length 502;

Best Local Similarity 21.99; Pred. No. 0.63;

Matches 110; Conservative 65; Mismatches 201; Indels 126; Gaps 26;

Qy 40 VSPVRNQNVTLPDGAWSAPPGSSAISNGKNPHOREPENLSAEAVTRGVGSMSTHWTC 99
| | | | | : : : : :
Db 74 VHLVIRNQARTPAAP-ATFASAPSS--NP----- 103
| | | | | : : : : :Qy 100 TPRIHP-----PMBSLPGICRP-KLSNDP-----AEDDKWENLYSEAE--RLIGTSTKE 146
| | | | | : : : : :
Db 104 TPSSOPNTNPFPAAGMGSPADILNPNDAWSMDNPITQQLGNPEEFKRTIQSNPQ 163
| | | | | : : : : :Qy 147 F-----DESTRHVL-----VLSQDAYKDR---QRIFRPLPLACHRLKNAPEVEHUSA 193
| | | | | : : : : :
Db 164 FOALIERNPEVGHILNDPNVNRQTMETIRNPNMEQEMMRNHQAIRNLQIGIP-----GG 217
| | | | | : : : : :Qy 194 ENLFHSIYNDKQKLFLLTNHRCRTRALRG-----GYEKKIGAAEVRNLLATR 243
| | | | | : : : : :
Db 218 EALERLYN-DVQPLNSATN-----SLSGNPFASLUGQSSEPRVDRAGQENNEALP 270
| | | | | : : : : :Qy 244 NPSSQLDSYIMAKVYVILASGAIGNPQILYNSGFSGLQVTPRNDLSLIPNLGRIITEQPMAF 303
| | | | | : : : : :
Db 271 NPWAS-----NANQATNQSNRSADFNLSLDSPGISLMQMSNPSNQASMF 319
| | | | | : : : : :Qy 304 CQIVLRFQEFVSDRDPDGLPWKEVAQHIKAPN--TDAL--PIPRDPPEQVTTPTFE 359
| | | | | : : : : :
Db 320 -----SPEVINSIR-----QNMNNPGLIDSIVGQIPSAARDNPQISEGIRR 360
| | | | | : : : : :Qy 360 EHPWHTOIHRDAFSYGAV-GPEVDSRVIVDLRFECATDPPEANNLL-VFQNDVQDGYSNPQ 417
| | | | | : : : : :
Db 361 SFQMLNMSDPSVMEAMRNPRVSEAFQIQEGFSTLRREAPQLLNLFQAGAMGGGAPGS 420
| | | | | : : : : :Qy 418 PTFRRYRSTASNVARKMKDMCEVASNIGLYLTPSPQFMDP--GLALHLAGTTRIGF- 474
| | | | | : : : : :
Db 421 DA-----NASSAGANSACLADLFN-SNMGGGRFSSTAAPVNPQTYASQLEQLQSMGFS 475
| | | | | : : : : :

Qy 475 DKATTVDNNLSLWDFANLYYA 496

Db 476 DRARNVA---ALTATFGDLNAA 494
| | | | | : | | | | |

RESULT 15

B70917

probable zwf2 protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

C:Accession: B70917

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987

A:Accession: B70917

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-514 <COL>

A:Cross-references: GB:Z95844; GB:AL123456; NID:G3250713; PIDN:CAB09259.1; PID:G21310

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: zwf2

C:Superfamily: glucose-6-phosphate dehydrogenase

Query Match 4.08; Score 116; DB 2; Length 514;

Best Local Similarity 22.89; Pred. No. 0.66;

Matches 98; Conservative 48; Mismatches 163; Indels 120; Gaps 23;

Qy 29 FVNIKGAQQ-VSPVRNQNVTLPDGAWSAPPGSSAISNGKNPHOREPENLSAEAVTR 87
| | | | | : | | | | |
Db 77 FGQVYVNAVQEHCRTPFQQNWDLAEG-FRFVGTGTFDDDDAFAQLAETLEKLDRE---R 132
| | | | | : | | | | |Qy 88 GVGMSHTWCTSTPRIHPM-ESL--PGIGRPKLSNDPAEDDKWENLYSEAEALIGTST 144
| | | | | : | | | | |
Db 133 GTGGNHAFYLAIPPKSPFVCEQLHKSLARPQ-----GDRWSRVVIE----- 175
| | | | | : | | | | |Qy 145 KEFDESIRHTLVLSQDAYKDRQIRFPLPLACHRL--KNAPEVVEHWSAENLPHSIYN 202
| | | | | : | | | | |
Db 176 KPGHGLASARELNKAVNAVPEEAVFR-----IDHLYGKETVQNTILARFANQLFDPDPIWN 231
| | | | | : | | | | |Qy 203 ----DDKOKKLFLLTNHRCRTRALTGGEKKIGAAE--VRN-----LLATRNPSOL 249
| | | | | : | | | | |
Db 232 AHYVDHVQ-----ITMAEDIGLGRAGYDYGIGAARDVIONHLMQALLATAMEEPVSFH 285
| | | | | : | | | | |Qy 250 DSYIMA-KVYVILASGAIGNP-----QILYNSGFSG----- 278
| | | | | : | | | | |
Db 286 PAALQAEKIKVLSATRLAEPLDQTTSRGOYAAAGWQGEKVVCLLDEEGFAEDSTTETFAA 345
| | | | | : | | | | |Qy 279 --LOVTPRNDSLIP-----NLGRYTEQPMAFQIVLRQEFV--DSVRDDPGLPWWK 327
| | | | | : | | | | |
Db 346 ITLEVDFRRWAGVFFYLRGTGKRLGRRVTEIALVF-----RRAPHLFPFDATMTDELG----- 396
| | | | | : | | | | |Qy 328 EAVAQHIKAPNPDALPIPRDPPEQVTTPTFEHPWHQIHRDA---PSYGAV-----GPE 380
| | | | | : | | | | |
Db 397 -----TNAMVIRVQ-PDEGVTLRFSGKVPGTANEVDRVNMDFSYGSAFAEDSPE 444
| | | | | : | | | | |Qy 381 VDSRVIVDL 389
| | | | | : | | | | |Db 445 AYERLILDV 453
| | | | | : | | | | |

Search completed: September 3, 2002, 15:58:48
Job time: 533 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 15:57:29 ; Search time 31.6 Seconds
(without alignments)
419.718 Million cell updates/sec

Title: US-09-856-327-2_COPY_76_618

Perfect score: 2901

Sequence: 1 NAEETAVPYVGYHKKNEI.....IINTLKGGTGKNTGEHRNL 543

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1142.5	39.4	623	1	US-08-734-925-2
2	1119	38.6	622	4	US-09-305-381-2
3	1024.5	35.3	586	4	US-09-023-731-1
4	192	6.6	51	4	US-09-023-731-12
5	159	5.5	615	4	US-09-297-937C-9
6	159	5.5	1276	4	US-09-297-937C-13
7	130.5	4.5	58	4	US-09-023-731-11
8	115	4.0	50	4	US-09-023-731-14
9	108	3.7	35	4	US-09-023-731-15
10	104.5	3.6	1876	2	US-08-609-049A-12
11	104.5	3.6	1876	4	US-09-170-996-12
12	102.5	3.5	61	4	US-09-023-731-6
13	96.5	3.3	1876	2	US-08-609-049A-28
14	96.5	3.3	1876	4	US-09-170-996-28
15	95.5	3.3	1257	1	US-08-049-783-2
16	95.5	3.3	1257	1	US-08-158-232-6
17	95.5	3.3	1257	1	US-08-304-626-6
18	95.5	3.3	1257	1	US-08-316-301A-6
19	95.5	3.3	1257	2	US-08-611-928-6
20	95.5	3.3	1257	3	US-09-173-891-6
21	95.5	3.3	1257	4	US-09-076-137-6
22	95.5	3.3	1257	5	PCT-US92-03624-6
23	95	3.3	3170	2	US-07-642-734C-5
24	95	3.3	3170	3	US-08-439-009A-5
25	92	3.2	2802	4	US-09-542-331-1
26	90.5	3.1	806	3	US-08-549-515-5
27	90.5	3.1	806	3	US-08-549-515-11

28	90	3.1	543	3	US-09-199-229-2	Sequence 2, Appli
29	90	3.1	543	4	US-09-443-087-2	Sequence 2, Appli
30	90	3.1	543	4	US-09-687-298-2	Sequence 2, Appli
31	90	3.1	807	4	US-08-974-549A-5	Sequence 5, Appli
32	90	3.1	1132	3	US-08-851-843A-225	Sequence 225, App
33	90	3.1	1132	4	US-08-974-549A-2	Sequence 2, Appli
34	90	3.1	1132	4	US-08-974-549A-344	Sequence 344, App
35	90	3.1	1132	4	US-08-854-050-225	Sequence 225, App
36	90	3.1	1132	4	US-09-430-323-225	Sequence 225, App
37	90	3.1	1132	4	US-09-128-354-2	Sequence 2, Appli
38	90	3.1	1154	4	US-08-974-549A-611	Sequence 611, App
39	90	3.1	1189	4	US-08-974-549A-613	Sequence 613, App
40	90	3.1	1200	4	US-08-974-549A-612	Sequence 612, App
41	90	3.1	1285	4	US-08-974-549A-600	Sequence 600, App
42	90	3.1	1407	4	US-08-974-549A-628	Sequence 628, App
43	89	3.1	514	4	US-08-974-549A-605	Sequence 605, App
44	88.5	3.1	1420	2	US-08-540-804-14	Sequence 14, Appl
45	88.5	3.1	1420	2	US-08-218-265-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-734-925-2
; Sequence 2, Application US/08734925
; Patent No. 5712139
; GENERAL INFORMATION:
; APPLICANT: NISHIMURA, IKUKO
; APPLICANT: OKADA, KIMIHARU
; APPLICANT: MINAMIHARA, TOMOYUKI
; APPLICANT: KAWAI, GENSUHIRO
; APPLICANT: KOYAMA, YASUJI
; APPLICANT: SUZUKI, MASARU
; TITLE OF INVENTION: NOVEL PYRANOSE OXIDASE, PYRANOSE OXIDASE
; TITLE OF INVENTION: GENE, NOVEL RECOMBINANT DNA AND PROCESS FOR PRODUCING
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESS: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/734,925
; FILING DATE: 22-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/568,428
; FILING DATE: 06-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 7126-002-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 623 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-734-925-2

```
Query Match      39.4%; Score 1142.5; DB 1; Length 623;
Best Local Similarity 45.1%; Pred. No. 3.4e-112;
Matches 251; Conservative 69; Mismatches 163; Indels 73; Gaps 15;

QY 13 GYHKNEIFQKIDRFVNVIGALQOVSVVVRNQNVPFLDPGANSAPGSSAISNGKNP 72
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 87 GAHKNTVEYQKIDKFVNVIOQLMSVSPVNTLVDTLSPSWQA--SSPFVRGNSP 144
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 73 HOREFENISAEAVTRGVGGMSTHTWCSTPRIHPPMESLPGIGRPKL-SNDPAEDDKENW 131
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 145 EODPLRLNSQAVTRVVGGMSTHTWCATPRFEQ-----RPLLVKDDQDADAENDR 197
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 132 LYSEARLIGTSTKFEDESIRHTLVLSLQDAYKDRQIRFRPLPLACHRLKNAPEVVEWH 191
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 198 LYTKAESYFKTCTDQKESIRHNVLNKLAEYKG-QRDFQIPLAATR--RSPTFEVWS 254
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 192 SAENLF--HSIYNDKQKLFLLTNHRCRLALTGGEYKKGAAEVRNLLATRNPSOL 249
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 255 SANTVFDLQNRPNTPDAPNERENLFFAVACERV-----VRN-----TSNBEI 295
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 250 DSY-----IMAKVYVLASGAINCPQILYNSGSGFL-QVTPRN-DSLIPNLGRY 295
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 296 ESLHLDLISGRFEKADVFVTAGAVHNAQLLVNSGFGQLGRDPDPANPPQLPSLSGY 355
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 296 ITEQPMACQIVLROEFVDSVRDDPY--GLP-----WVKEAVAQH 333
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 356 ITEQSLVFCQVMTSELIDSVKSMILRGNGDLGYSVYTPGAEKNKHPDMWNEKVKNH 415
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 334 IAKNTDALPIPRDPEQVTPTEEHHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRWFG 393
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 416 MNQHQEDPLPIPEPEQVTTLPQSPHWPHTQIHRDAFSYGAVQGSIDSLRILVDWRFRG 475
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 394 ATDPANLLVQNDVODGYSMPQTPRYR-PTASNVARKMMADMCEVASNLGYLPT 452
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 476 RETPEENKLFESDKITDYNPQTFDFRFPAGRTSKAEADMTDMCMVSAKIGGFLLPG 535
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 453 SPQPMDFGLALHLAGTTRIGFDKA--TTVADNNSLVDFANLYVAGNCTIRTGFCENPT 510
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 536 SLQPMEPLGLVHLGGTHRMGFEDEQKCCVNTDSRVFGKFLGCGGNIPTAYGANPT 595
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 511 LTSMCHAKTSARSIIN 526
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 596 LTAMSLATKSCYEIKN 611
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 2
US-09-305-381-2
; Sequence 2, Application US/09305381
; Patent No. 6146865
; GENERAL INFORMATION:
; APPLICANT: Christensen, Soren
; APPLICANT: Lassen, Soren Flensted
; APPLICANT: Schneider, Fale
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; TITLE OF INVENTION: Having Pyranose Oxidase Activity
; FILE REFERENCE: 5571.200-US
; CURRENT APPLICATION NUMBER: US/09/305.381
; EARLIER FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/088,724
; EARLIER FILING DATE: 1998-06-10
; EARLIER APPLICATION NUMBER: PA 1998 00774
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 622
; TYPE: PRT
; ORGANISM: Corioliolus Versicolor
US-09-305-381-2

Query Match      38.6%; Score 1119; DB 4; Length 622;
Best Local Similarity 44.6%; Pred. No. 1.1e-109;
Matches 251; Conservative 69; Mismatches 183; Indels 60; Gaps 15;

QY 13 GYHKNEIFQKIDRFVNVIGALQOVSVVVRNQNVPFLDPGANSAPGSSAISNGKNP 72
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 87 GSHKNTVEYQKIDKFVNVIOQLMPVSPVNTVYDVLSPASWQA--STFFVRNGANP 144
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 73 HOREFENISAEAVTRGVGGMSTHTWCSTPRIHPPMESLPGIGRPKL-SNDPAEDDKENW 131
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 145 EODPLRLNSQAVTRVVGGMSTHTWCATPRFEK-----LQRLVLKNDKADDAENDR 197
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 132 LYSEARLIGTSTKFEDESIRHTLVLSLQDAYKDRQIRFRPLPLACHRLKNAPEVVEWH 191
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 198 LYTKAESYFKTCTDQKESIRHNVLNKLAEYKG-VRDFQIPLAATR--QSPTEVWS 254
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 192 SAENLF--HSIYNDKQKLFLLTNHRCRLALTGGEYKKGAAEVRNLLATRNPSOL 249
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 255 SAHTVFDLENRENPKDAPKQRFNLFPAVACTNVRNDRNANSEIVG-LDVRDLHGKSKIT--- 310
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 250 DSYIMAKVYVLASGAINCPQILYNSGSGFLQVTPRND-----SLIPNLGRYITEQPMAP 303
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 311 ---IRAKVITLTAGAVHNAQLLAASGFGQL---GRDPDAKPLPSLLPYLGTHITEQLVIF 364
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 304 COIVLRQEFVDSVRDDP--YGLP-----WVKEAVAQHAKNPTDA 341
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 365 CQTVAMSTELINSVTADMTIVGKPGHPDYSVTYTPGNPNKNKHPDMWNEKVKKMMHQEDP 424
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 342 LPIPRDPEQVTPTEEHHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRWFGATDPEANN 401
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 425 LPIPRDPEQVTTLPQATHPWHTQIHRDAFSYGAVQGSIDSLRILVDWRFRGTEPKEEN 484
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 402 LLVFONDVODGYSMPQTPRYR-PTASNVARKMMADMCEVASNLGYLPTSPPOFMDPG 461
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 485 KLWFSKIDTAYNLQPTDFR--FPGREAEDMTDMCMVSAKIGGFLLPGYPMFEPG 542
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 462 LALHLAGTTRIGFDKAT--TVADNNSLVDFANLYVAGNCTIRTGFCENPTLTSCHAIK 519
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 543 LVHLGGTHRMGFEKADKCCVNTDSRVFGKFLGCGGNIPTAYAAANPTLTAMSLATK 602
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 520 SANSIINTLKGTDGKNTGEHEN 542
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 603 SCEYI---KKNEPSPNPVKHHN 622
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 3
US-09-023-731-1
; Sequence 1, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akihiko; Izumo, Koji; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
```


REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 58
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-023-731-11

Query Match 4.5%; Score 130.5; DB 4; Length 58;
Best Local Similarity 44.4%; Pred. No. 1.5e-06;
Matches 28; Conservative 11; Mismatches 13; Indels 11; Gaps 2;
QY 258 YVLASGAIGNPQILYNSGSLQVTPRNDLSLIPNLGRYTEQPMACQIVLRQEFVDSVR 317
DB 2 FVIAAGAVCTPQILWNS-----NIRPY-----ALGRYLSQSQMTFCQIVLKRGIYDAIA 50
QY 318 DDP 320
DB 51 TDP 53

RESULT 8
US-09-023-731-14
Sequence 14, Application US/09023731
Patent No. 6291648
GENERAL INFORMATION:
APPLICANT: Kawamura, Yukio; Morita,
APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.
TITLE OF INVENTION: ANTI-TUMOR PROTEIN AND
TITLE OF INVENTION: GENE ENCODING SAME
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,731
FILING DATE: 13-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 50
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-023-731-14

Query Match 4.0%; Score 115; DB 4; Length 50;
Best Local Similarity 53.5%; Pred. No. 5.2e-05;
Matches 23; Conservative 5; Mismatches 15; Indels 0; Gaps 0;
QY 409 VODGYSPQPTFRYPSTASNVARKKMDMCEVASNLGGYLP 451
DB 8 VTDYIGMPQPTFFHKRTNADGDRDQRMMNDMTNVANMLGGYLP 50
RESULT 9
US-09-023-731-15
Sequence 15, Application US/09023731
Patent No. 6291648
GENERAL INFORMATION:
APPLICANT: Kawamura, Yukio; Morita,
APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.
TITLE OF INVENTION: ANTI-TUMOR PROTEIN AND
TITLE OF INVENTION: GENE ENCODING SAME
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,731
FILING DATE: 13-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792

INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 35
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-023-731-15

Query Match 3.7%; Score 108; DB 4; Length 35;
Best Local Similarity 66.7%; Pred. No. 0.00015;
Matches 22; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
QY 453 SPQFMDPGLALHLAGTTRIGFDKATTVADNNS 485
DB 2 SYQFMAPGLVLHITGTRIGTDDQTSVADPTS 34

RESULT 10
US-08-609-049A-12
Sequence 12, Application US/08609049A
Patent No. 5948664
GENERAL INFORMATION:
APPLICANT: Williams, Lewis T.
APPLICANT: Molz, Lisa
APPLICANT: Chen, Yen-Wen
TITLE OF INVENTION: No. 5948664el PI 3-Kinase Polypeptides
NUMBER OF SEQUENCES: 32

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Townsend and Townsend and Crew LLP
;; STREET: Two Embarcadero Center, 8th Floor
;; CITY: San Francisco
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94111-3834
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/609,049A
;; FILING DATE: 29-FEB-1996
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Dow, Karen B.
;; REGISTRATION NUMBER: 29,684
;; REFERENCE/DOCKET NUMBER: 2307K-063700US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415-326-2400
;; TELEFAX: 415-326-2422
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1876 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-609-049A-12

Query Match 3.6%; Score 104.5; DB 2; Length 1876;
Best Local Similarity 22.0%; Pred. No. 0.37;
Matches 85; Conservative 53; Mismatches 170; Indels 79; Gaps 21;
QY 8 VPYVGYHKKNEIEFKQDIDREVNVIKALQOVSVVPRNQNVTLPDGA---WSAP--P 61
DB 184 VPYQAAQQOORPLN--SEELQRLYSN---PAQNAVVPVQPNAYMYYPGAVVTPYTAIYP 239
QY 62 GSSAISNKNPHQREFENLSAEAVTRGVGGMSTHTCTTPRIHP-PMESLPGIG--RPKL 118
DB 240 GSAFMPFPQYPAQ-----GYGFGGAYTHMDLRRPQSQAPAQQTATTTSHHSQP 288
QY 119 SNDPAEDDKENNELYSEAE-----LIGTSTKEFDESIRHTLVRLSLQDA---YKDRORI 170
DB 289 SNHSTSSPAEANGVAFAPARRQVPSTVGSSSHTGNNGHSSVPRRGNDLIDLNHEDYSRV 348
QY 171 -----FRPLPLACHRLKNAPEYVEVHSAENLFHSIYNDDKQKKLFTLLTNHRCRTRALGTG 225
DB 349 SVLEAFDPLNDNTGNDTASDSTSYAEVDPDFLYSGD-----AATQYSDPMYEA 401
QY 226 GYEKKIGAEVNR-NLLATR-----NPSQOLDYSYMA-----KYVVLASGAIGNP----- 268
DB 402 RWDKTAVTVPNVGLIGRWQDFLSQPSSTSSQYGVAPPEESLKLAEANGSGTISPPPLPP 461
QY 269 --QILYNSGFGLOVT-PRNDSLIPNLGRYTEQPMACQIVLRO-----EFVDSVR 317
DB 462 RNQCYESNQAMPVSRPPQSSVLTD--SYTSSIP---ANVVLDRKCTCTRLYELISDQR 516
QY 318 -DDPYGLPWKKEAVAQHIKNTDPL 343
DB 517 TDDPELLEFY-HMWKEVARYPHDDAP 542

RESULT 11
US-09-170-996-12
; Sequence 12, Application US/09170996
; Patent No. 6291220
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Moliz, Lisa

;; APPLICANT: Chen, Yen-Wen
;; TITLE OF INVENTION: No. 6291220el PI 3-Kinase Polypeptides
;; NUMBER OF SEQUENCES: 32
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Townsend and Townsend and Crew LLP
;; STREET: Two Embarcadero Center, 8th Floor
;; CITY: San Francisco
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94111-3834
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/170,996
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/609,049
;; FILING DATE: 29-FEB-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Dow, Karen B.
;; REGISTRATION NUMBER: 29,684
;; REFERENCE/DOCKET NUMBER: 2307K-063700US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415-326-2400
;; TELEFAX: 415-326-2422
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1876 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-09-170-996-12

Query Match 3.6%; Score 104.5; DB 4; Length 1876;
Best Local Similarity 22.0%; Pred. No. 0.37;
Matches 85; Conservative 53; Mismatches 170; Indels 79; Gaps 21;
QY 8 VPYVGYHKKNEIEFKQDIDREVNVIKALQOVSVVPRNQNVTLPDGA---WSAP--P 61
DB 184 VPYQAAQQOORPLN--SEELQRLYSN---PAQNAVVPVQPNAYMYYPGAVVTPYTAIYP 239
QY 62 GSSAISNKNPHQREFENLSAEAVTRGVGGMSTHTCTTPRIHP-PMESLPGIG--RPKL 118
DB 240 GSAFMPFPQYPAQ-----GYGFGGAYTHMDLRRPQSQAPAQQTATTTSHHSQP 288
QY 119 SNDPAEDDKENNELYSEAE-----LIGTSTKEFDESIRHTLVRLSLQDA---YKDRORI 170
DB 289 SNHSTSSPAEANGVAFAPARRQVPSTVGSSSHTGNNGHSSVPRRGNDLIDLNHEDYSRV 348
QY 171 -----FRPLPLACHRLKNAPEYVEVHSAENLFHSIYNDDKQKKLFTLLTNHRCRTRALGTG 225
DB 349 SVLEAFDPLNDNTGNDTASDSTSYAEVDPDFLYSGD-----AATQYSDPMYEA 401
QY 226 GYEKKIGAEVNR-NLLATR-----NPSQOLDYSYMA-----KYVVLASGAIGNP----- 268
DB 402 RWDKTAVTVPNVGLIGRWQDFLSQPSSTSSQYGVAPPEESLKLAEANGSGTISPPPLPP 461
QY 269 --QILYNSGFGLOVT-PRNDSLIPNLGRYTEQPMACQIVLRO-----EFVDSVR 317
DB 462 RNQCYESNQAMPVSRPPQSSVLTD--SYTSSIP---ANVVLDRKCTCTRLYELISDQR 516
QY 318 -DDPYGLPWKKEAVAQHIKNTDPL 343
DB 517 TDDPELLEFY-HMWKEVARYPHDDAP 542

RESULT 12

```

US-09-023-731-6
; Sequence 6, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akhiho; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-023-731-6

Query Match 3.5%; Score 102.5; DB 4; Length 61;
Best Local Similarity 43.5%; Pred. No. 0.0016;
Matches 27; Conservative 6; Mismatches 24; Indels 5; Gaps 2;

QY 29 FVNIWIGALQVSVVVRNQVPTLDPGANSAPGSSA-----ISNGKNPHOREFENISAE 84
|||:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 FVNIINGALQPISTPSDVTQPTLAVAAW-APPIDPAEGQLVIMGHPNQGAINLP 59
QY 85 VT 86
||
Db 60 VT 61

RESULT 13
US-08-609-049A-28
; Sequence 28, Application US/08609049A
; Patent No. 5948664
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 5948664el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA

US-09-023-731-6
; Sequence 6, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akhiho; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-023-731-6

Query Match 3.5%; Score 102.5; DB 4; Length 61;
Best Local Similarity 43.5%; Pred. No. 0.0016;
Matches 27; Conservative 6; Mismatches 24; Indels 5; Gaps 2;

QY 29 FVNIWIGALQVSVVVRNQVPTLDPGANSAPGSSA-----ISNGKNPHOREFENISAE 84
|||:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 FVNIINGALQPISTPSDVTQPTLAVAAW-APPIDPAEGQLVIMGHPNQGAINLP 59
QY 85 VT 86
||
Db 60 VT 61

RESULT 13
US-08-609-049A-28
; Sequence 28, Application US/08609049A
; Patent No. 5948664
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 5948664el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA

US-09-023-731-6
; Sequence 6, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akhiho; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-023-731-6

Query Match 3.5%; Score 102.5; DB 4; Length 61;
Best Local Similarity 43.5%; Pred. No. 0.0016;
Matches 27; Conservative 6; Mismatches 24; Indels 5; Gaps 2;

QY 29 FVNIWIGALQVSVVVRNQVPTLDPGANSAPGSSA-----ISNGKNPHOREFENISAE 84
|||:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 FVNIINGALQPISTPSDVTQPTLAVAAW-APPIDPAEGQLVIMGHPNQGAINLP 59
QY 85 VT 86
||
Db 60 VT 61

RESULT 13
US-08-609-049A-28
; Sequence 28, Application US/08609049A
; Patent No. 5948664
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 5948664el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA

US-09-023-731-6
; Sequence 6, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akhiho; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-023-731-6

Query Match 3.5%; Score 102.5; DB 4; Length 61;
Best Local Similarity 43.5%; Pred. No. 0.0016;
Matches 27; Conservative 6; Mismatches 24; Indels 5; Gaps 2;

QY 29 FVNIWIGALQVSVVVRNQVPTLDPGANSAPGSSA-----ISNGKNPHOREFENISAE 84
|||:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 FVNIINGALQPISTPSDVTQPTLAVAAW-APPIDPAEGQLVIMGHPNQGAINLP 59
QY 85 VT 86
||
Db 60 VT 61

RESULT 13
US-08-609-049A-28
; Sequence 28, Application US/08609049A
; Patent No. 5948664
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 5948664el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA

US-09-023-731-6
; Sequence 6, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akhiho; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-023-731-6

Query Match 3.5%; Score 102.5; DB 4; Length 61;
Best Local Similarity 43.5%; Pred. No. 0.0016;
Matches 27; Conservative 6; Mismatches 2
```

```

: STATE: California
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/170,996
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/609,049
: FILING DATE: 29-FEB-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Dow, Karen B.
: REGISTRATION NUMBER: 29,684
: REFERENCE/DOCKET NUMBER: 2307K-063700US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-326-2400
: TELEFAX: 415-326-2422
: INFORMATION FOR SEQ ID NO: 28:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1876 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-170-996-28

Query Match 3.3%; Score 96.5; DB 4; Length 1876;
Best Local Similarity 21.7%; Pred. No.2.6;
Matches 84; Conservative 53; Mismatches 171; Indels 79; Gaps 21;

Qy 8 VYYVPGTHKKNEIEFQKIDIRFNVVIGALQQVSVPRVQNWNPTLDPCA----WSAP--P 61
||| | : : : ||| | ||| | ||| :||| |
Db 184 VPYQAAQQQQLN--SEELQRLYSN---PAQMAVVPVPQPNAYMYPGAVVTPYTATPVP 239
||| | : : : ||| | ||| | ||| :||| |
Qy 62 GSSAISNGKNPHQREFENLSAEATRGVGGNSTHWCTSPRIHP--PMESLPGIG--RPKL 118
||| | : : : ||| | ||| | ||| :||| |
Db 240 GSAAFMPQPPYPAQ-----GYGFGGAYTHMDLRRPOSQAPAQQTATTTSHHSQP 288
||| | : : : ||| | ||| | ||| :||| |
Qy 119 SNDAPEDDKENNELYSAEAR----LIGTSTKEFDESIRHTLVLRSLQDA----YKDRQRI 170
||| | : : : ||| | ||| | ||| :||| |
Db 289 SNHSTSSPAEANGVAFPARRQVPSTGVGSSSSHTGNGCHSVPRRGNDLIDLNHEDYSRV 348
||| | : : : ||| | ||| | ||| :||| |
Qy 171 -----FRPLPLACHRLKNAPYVZVHSAENLFHSIYNDKOKKLETLTNRCTRRLAUG 225
||| | : : : ||| | ||| | ||| :||| |
Db 349 SVLEAFQPLNDNTGNTASDTSYIAYEYDFPFDLYSGD-----AATQYSDPMYEAVN 401
||| | : : : ||| | ||| | ||| :||| |
Qy 226 GYEKKIGAAEVR--NLATR-----NPSSQLDSYIWA-----KVVVLASGAIGNP----- 268
||| | : : : ||| | ||| | ||| :||| |
Db 402 RWDKTVATVSNVGLIGWRQDFLQSPSTSSSQYGVAPPEESLKAENGSETISPPPLPP 461
||| | : : : ||| | ||| | ||| :||| |
Qy 269 --QILYNSGFSGLQVT--PRNDSLIPNLGRYITEQPMFACQIVLRQ-----EFVDSVR 317
||| | : : : ||| | ||| | ||| :||| |
Db 462 RNOQCYESNQAAVPRPPQSSVLTD--SYTSSIP---ANVLDRRKTCTRLYELISDQR 516
||| | : : : ||| | ||| | ||| :||| |
Qy 318 -DDPYGLPWWKEAVAQHIKNPTDALP 343
||| | : : : ||| | ||| | ||| :||| |
Db 517 TDDPELLEFY-HMVKEVRARYPHDDAP 542
||| | : : : ||| | ||| | ||| :||| |

RESULT 15
US-08-049-783-2
: Sequence 2, Application US/08049783
: Patent No. 5439881
: GENERAL INFORMATION:
: APPLICANT: Narva, Kenneth E
: APPLICANT: Schwab, George E
: APPLICANT: Payne, Jewel M
: TITLE OF INVENTION: Gene Encoding No. 5439881el Nematode-Active

```

```

; TITLE OF INVENTION:   Toxins Cloned from Bacillus thuringiensis Isolates
;
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jeff Lloyd
; STREET: 2421 N.W. 41st Street
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/049,783
; FILING DATE: 19930419
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lloyd, Jeff
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 2 (PS33F2):
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1257 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; INDIVIDUAL ISOLATE: PS33F2
; IMMEDIATE SOURCE:
; CLONE: PS33F2a
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..1257
;
US-08-049-783-2

Query Match          3.3%; Score 95.5; DB 1; Length 1257;
Best Local Similarity 20.7%; Pred. No. 1.6;
Matches 120; Conservative 78; Mismatches 206; Indels 175; Gaps

Qy      12 PGYHKKEIEFKDIDRFVNVTKGALQOVSVPRNQNVPT-LDPGAW SAPPGSSAISNGK 70
        | : :: : : : : : : : : : : : : : : : : : : : : : : : : :
Db      359 PNTHETHYHDFSYNTQSGNISRGSSNPIDILNPIISTCIRNSFYKATAGS VLVNEK 418
        | : :: : : : : : : : : : : : : : : : : : : : : : : : : :

Qy      71 NPQHOREFNLSAEAVTRGVGGSHTWCTSPRIHPPELPGICRPKLSDPAEDDKENN 130
        | : || : || : || : || : || : || : || : || : || : || : || : ||
Db      419 DGTQ-----GYAFQAAPTGA-----WD-----HSFIES-----DGAPGHKLN 452

Qy      131 ELYSEARELIGSTKEFDESISRHTLV-LRSLDQAYKQRIFRLPLACHLKN----- 183
        | : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      453 YIYTSP----GDTRLDFINV--YTLISTPTINELSTEKIGF---PAEKGYIKNQIMKY 503

Qy      184 --APEYVEVHSAENLFHSYINDDKOKFLFTLLTNHRCTRIALTGGYEKKTGAAEVRNL-- 239
        ||| : || : || : || : || : || : || : || : || : || : || : ||
Db      504 YGKPEYINGAQPVNL-----ENQOTLIFE-----HASKTAQYITIRIVASTQTCKG 550

Qy      240 -----LATRNPPSQLDSYIMAKVVYVLASCAIGNPOLIYN SGFSGLQVTPRNDSLIPN 291
        | : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      551 YFRLDNQELQTLNIPTSHNGYV-----TCNIGENYDLTYTIG--SYTTIEGNHTL--- 597

Qy      292 LGRYTEQPMAFCQI-----VL-RQEFV--DSVRDDPYGLPWKKEAVAQHI AKNPTD 340
        |||| : || : || : || : || : || : || : || : || : || : || : ||
Db      598 -----QIQHNDKNCMVLDRIEFVPKLSQDSP-----HATQIRDAFSYGAVGPEVDSRVIVDL 389
        | : || : || : || : || : || : || : || : || : || : || : || : ||

```

RESULT 15

US-08-049-783-2

; Sequence 2, Application US/08049783

; Patent No. 5439881

; GENERAL INFORMATION:

APPLICANT: Narva, Kenneth E

APPLICANT: SCHWAB, George E

APPLICANT: PAYNE, JEWEL M
TITLE OF INVENTION: Gene Encoding No. 5439881e1 Nematode-Active

Db 625 -----QDSPPEVHESTIIIDKSSPTIWSNKHSHYSHIHLEG-SYTSQG-SYPHNLLINL 676

Qy 390 RWFQATDPEANNLLVFQF-DVQDGYSMPOPTFRYRPSSTASNVRRKMMADMCEVASNLG- 447

Db 677 --FHPTDPRNRNHTIHVWNGDMNVYD-----KDSVADCLNFENKITATIPSDAWYSGT 726

Qy 448 -----GYLPSPQPFMDPGLALHLAG-TTRIGFDKAT-----TVADNNSLYWDFANL 493

Db 727 ITSMHLEFNDNFKTITPKF---ELSNLENIITTQVNALFASSAQDTILASNVSDYW-IEQV 782

Qy 494 YVAGNGTIRTGFG-ENPTLTSMCHAISARSIIINTLKGG 531

Db 783 VMKYDALSDEVFGEKKALRKLNVQAKRLSKIRNLLIGG 821

Search completed: September 3, 2002, 15:57:34

Job time: 3144 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 15:56:34 ; Search time 82.18 Seconds
(without alignments)
733.914 Million cell updates/sec

Title: US-09-856-327-2_COPY_76_618

Perfect score: 2901

Sequence: 1 NAEETAVPYGPKHKEI.....INTLKGTDCKNTGHRNL 543

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :	A_Geneseq_032802.*
1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2901	100.0	618	22	Lyophyllum shimeji
2	1142.5	39.4	623	20	Coriolus versicolor
3	1136	39.2	633	21	Pleurotus cornucop
4	1128.5	38.9	623	17	Pyranose oxidase
5	1119	38.6	622	22	Trametes hirsuta p
6	1024.5	35.3	566	20	An antitumour prot
7	976.5	33.7	564	21	T. matsutake pyran
8	976.5	33.7	564	21	Trichoderma deri
9	192	6.6	51	20	Peptide derived fr
10	177	6.1	67	21	T. matsutake pyran
11	177	6.1	67	21	N-terminal fragmen

12	168	5.8	34	22	AA970334	Lyophyllum shimeji
13	162	5.6	30	22	AA970331	Lyophyllum shimeji
14	159	5.5	615	20	AA905719	Glucanase dehydrog
15	157	5.4	544	20	AA914049	G. oxydans D-sorbi
16	142	4.9	29	22	AA970333	Lyophyllum shimeji
17	130.5	4.5	58	20	AA987540	Peptide derived fr
18	128.5	4.4	657	21	AA915525	Arabidopsis thalia
19	128.5	4.4	675	21	AA915524	Arabidopsis thalia
20	128.5	4.4	748	21	AA915523	Arabidopsis thalia
21	121	4.2	24	22	AA970332	Lyophyllum shimeji
22	115	4.0	50	20	AA987543	Peptide derived fr
23	114.5	3.9	704	20	AA931754	Candida tropicalis
24	110	3.8	551	21	AA913598	Streptomyces globi
25	108	3.7	35	20	AA987544	Peptide derived fr
26	106.5	3.7	778	20	AA935090	Chlamydia pneumoni
27	106	3.6	546	20	AA928581	Choline oxidase (C
28	105	3.6	609	22	AA956302	Propionibacterium
29	105	3.6	833	22	AB968116	Drosophila melanog
30	104.5	3.6	1356	22	AB967291	Drosophila melanog
31	104.5	3.6	1876	18	AA938757	Phosphatidyl inosi
32	104.5	3.6	1876	22	AB961970	Drosophila melanog
33	104.5	3.6	1876	22	AB966876	Drosophila melanog
34	103	3.6	437	22	AB929360	Novel human diagno
35	103	3.6	1688	22	AB959327	Drosophila melanog
36	102.5	3.5	61	20	AA987535	Peptide derived fr
37	102.5	3.5	352	22	AB971452	Drosophila melanog
38	102.5	3.5	1686	19	AA970991	Human class II P13
39	102	3.5	1130	20	AA913446	Human aortic carbo
40	102	3.5	1244	22	AB965766	Human protein kina
41	100.5	3.5	880	22	AB971719	Drosophila melanog
42	100	3.4	522	22	AB958468	Drosophila melanog
43	100	3.4	522	22	AB958474	Drosophila melanog
44	100	3.4	645	21	AA941556	Human ORFX ORF1320
45	98.5	3.4	777	22	AB919089	Novel human diagno

ALIGNMENTS

RESULT 1

AA970335

ID AA970335 standard; Protein; 618 AA.

XX AC AA970335;

DT 20-JUL-2001 (first entry)

XX Lyophyllum shimeji antibacterial protein.

XX Fungus; antibacterial; antibiotic; plant pathogen; bacteri

XX Pyricularia orizae; Rhizoctonia solani; rice pathogen

XX Lyophyllum shimeji.

XX WO200121657-A1.

XX 29-MAR-2001.

XX 20-SEP-2000; 2000WO-JP06404.

XX 21-SEP-1999; 99JP-0267238.

XX (NISB) JAPAN TOBACCO INC.

XX (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.

XX Takakura Y, Kuwata S, Inoue Y;

XX WPI; 2001-281548/29.

XX N-PSDB; AAF99980.

XX Antibacterial protein and encoded gene isolated from Lyophyllum

XX shimeji, with activity against plant pathogenic bacteria, applicable in

XX agriculture e.g. rice cultivation at low concentration, produced at low

```

PT cost on large scale
XX Claim 3; Page 43-45; 52pp; Japanese.
XX
XX The present sequence is an antibacterial protein from the fungus
CC Lyophyllum shimeji. The protein was obtained from a fraction prepared
CC by extracting lyophyllum shimeji with water and subtracting the extract
CC to ammonium sulphate precipitation. The protein inhibits the growth
CC of the plant pathogenic bacteria Pyricularia oryzae and Rhizoctonia
CC solani at a relatively low concentration. P. oryzae and R. solani are
CC causative of the two major diseases of rice. The protein contains
CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.
XX The antibacterial protein can be produced at low cost on a large scale.
XX
SQ Sequence 618 AA;

Query Match 100.0%; Score 2901; DB 22; Length 618;
Best Local Similarity 100.0%; Pred. No. 5.1e-266;
Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NAEGETAVPYPCYHKKEIEFKDIDRFVNVKIGALQOVSVPRNQNVTLDPGAWSAP 60
   |||||
Db 76 naeegtavpyvpyghknieiefqkdidrfvnhkgaigqvsvprnqnvtldpgawsap 135
   |||||
QY 61 PGSSAISNGKNPHOREFENLSAEAVTRGVGGMSTHTWCSTPRIHPPMESLPGIGRPKLSN 120
   |||||
Db 136 pyssaisngknphqrefenlsaeavtrgvvgmsthtwcstprihppmeslpgigrpklsn 195
   |||||
QY 121 DPAEDDKENNELYSEARLIGSTKEFDESIRHTLVRLSLQDAYKDRQIRFRLPLACHR 180
   |||||
Db 196 dpaeddkewnelysearligstkefdesirhtlvrlslqdaykdrqirfrlplachr 255
   |||||
QY 181 LKNAPYVWHSANLPHSIYNDKQKLFLLTNHRCITRLALTGGEKKIGAAEVRNLL 240
   |||||
Db 256 lknapevewhsaenlfhsinynddkqkllflltnhrcitrlaltggyekkgaaevrnll 315
   |||||
QY 241 ATRNPSSQLDSYIMAKVYVLSAGAIQNPQILYNSGFSGLQVTPRNDLSLIPNLGRYITEQ 300
   |||||
Db 316 atrnpssqldsyimakvyvlsagainpqilynsfsglqvtpnrndslipnlgryteqp 375
   |||||
QY 301 MAFCQIVLRQEVDSVRDPPYGLPWKKEAVAOHIAKNPTDALPIPRDPPEQVTPPTTEE 360
   |||||
Db 376 mafcqvrlrqetvdsrvddpyglpwkavachiaaknptdalpirdpdpqvtpptfee 435
   |||||
QY 361 HPWHTQIHRDAFSYGAVGPEVDSRVLDLRFWFGATDPEANNLLVFQNDVQDGYSMPOPTF 420
   |||||
Db 436 hpwhtqihrdafsygavgpevdsrvldlrfwfgatdpeannllvfqndvqdgysmpoptf 495
   |||||
QY 421 RYRPTASTNVRARKMADMCCEVASNLGGYLPSPQFMDPGLALHLAGTTRIGFDKATTV 480
   |||||
Db 496 ryrptastnvrarkmadmcevasnlggylpqfmdpglalhlagttrigfdkattv 555
   |||||
QY 481 ADNNSLVWDFANLYVAGNTIRTGGENPTLTSCHAKSARSIINTLKGTDGKNTGEH 540
   |||||
Db 556 adnnslvwdfanlyvagnrtirtggenptltsmchaiksarsintlkgtdgkntgeh 615
   |||||
QY 541 RNL 543
   |||
Db 616 rnl 618

RESULT 2
ID AAW94308
XX AAW94308 standard; protein; 623 AA.
AC AAW94308;
XX
XX 08-APR-1999 (first entry)
XX
XX Coriolus versicolor pyranose oxidase.
DE
XX Coriolus versicolor; modified; pyranose oxidase; heat stability;

```

```

KW pH stability; glucose; glucosone.
XX Coriolus versicolor.
XX
XX JPI1009271-A.
XX
XX 19-JAN-1999.
XX
XX 24-JUN-1997; 97JP-0181865.
XX
XX 24-JUN-1997; 97JP-0181865.
XX
XX (KIKK ) KIKKOMAN CORP.
XX
XX WPI; 1999-145891/13.
XX
XX New modified pyranose oxidase gene and protein - useful for
PT efficient recombinant production of pyranose oxidase with excellent
PT Km, heat and pH stability
XX
XX Claim 1; Page 7-9; 10pp; Japanese.
XX
XX The present sequence represents pyranose oxidase derived from Coriolus
CC versicolor, which can be modified by adding, deleting or replacing at
CC least one amino acid (preferably where the amino acid at position 542
CC is replaced) and still retains pyranose oxidase activity. The modified
CC pyranose oxidase with the amino acid at position 542 replaced, has the
CC following properties: (a) Action: It oxidises glucose to glucosone;
CC (b) Substrate specificity: It reacts specifically with glucose and also
CC reacts with galactose, L-sorbose, D-xylose and L,5-anhydro-D-glucitol;
CC (c) Stable pH: 3.5 to 11.0 at 50 degrees Celsius for 30 minutes; (d)
CC Optimum pH: Near 6.5; (e) Optimum temperature: Near 55 degrees Celsius;
CC and (f) Temperature stability: Stable up to 55 degrees Celsius. The
CC present invention also describes a method for the preparation of a
CC modified pyranose oxidase. The method efficiently prepares a modified
CC pyranose oxidase with excellent Km value, heat stability and
CC pH stability.
XX
XX Sequence 623 AA;

Query Match 39.4%; Score 1142.5; DB 20; Length 623;
Best Local Similarity 45.1%; Pred. No. 4.1e-99;
Matches 251; Conservative 69; Mismatches 163; Indels 73; Gaps 15;

QY 13 GYHKKEIEFKDIDRFVNVKIGALQOVSVPRNQNVTLDPGAWSAPGSSAISNGRNP 72
   |||||
Db 87 gahkktvveyqknidkfnvvgqgimsvsvpvntlvdtltsptswga--ssffvngsnp 144
   |||
QY 73 HOREFENLSAEAVTRGVGGMSTHTWCSTPRIHPPMESLPGIGRPKLSNDPAE--DKEWNE 131
   |||
Db 145 eqdplrnlsqgavtrvggmsthtwcattrfdreq-----rpllvkddqdaadaewdr 197
   |||
QY 132 LYSEARLIGSTKEFDESIRHTLVRLSLQDAYKDRQIRFRLPLACHRLKNAPYVEWH 191
   |||||
Db 198 lytkaesyfkgtddqfkesirhnlvlnklaeeykg-qrdfigqlaatr--rsptfvevs 254
   |||
QY 192 SAENLF--HSIYNDKQKLFLLTNHRCITRLALTGGEKKIGAAEVRNLLATRPSSQL 249
   |||
Db 255 santvfdlqnrntdpapnerfnlpavacerv-----vrn-----tsnsei 295
   |||
QY 250 DSY-----IMAKVYVLSAGAIQNPQILYNSGFSGL-QVTPRN--DSLIPNLGRY 295
   |||
Db 296 eslhhdliisgdrfeikadvlvtagavhnaqlvnsfgqlgrpdpanppqlpslgsy 355
   |||
QY 296 ITEQPMACQIVLRQEVDSVRDDPY--GLP-----WKEAVAOH 333
   |||||
Db 356 iteqslvfcqvmstelidsvksdmilrnpqdlgysvtytpgaetnkhpdwnnekvnkh 415
   |||
QY 334 IAKNPTDALPIPRDPPEQVTPPTTEEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRFWG 393
   |||||
Db 416 mmqhgedplpfpfedpepqvttlflqpsbpwhtqlhrdaifsygavqgsidsrlivdwrffg 475

```


QY 394 ATDPEANLLVFQNDVODGYSMFQPTPRYR--PSTASNVARKMADMCCEVASNIGGYLPT 452
 Db 476 rtepkeenklwfsdkitdympqptfdfrpagrtskaedmntdmcvmsaklglfpg 535
 QY 453 SPQFMDPGALHLAGTTRIGFQKA--TTVDANNSLVWDFANLYVAGNGTIRTGFGENPT 510
 Db 536 slqfmepegvlhlggchrmgfdedkccvntdsrvfgkniflgcgcnlptayganpt 595
 QY 511 LTSMCHAIKSARSIIIN 526
 Db 596 ltamslaiksczeylkn 611

RESULT 3
 AAY52700
 ID AAY52700 standard; Protein; 633 AA.
 XX
 AC AAY52700;
 XX
 DT 07-MAR-2000 (first entry)
 XX
 DE Pleurotus cornucopiae antitumour protein.
 XX
 KW Antitumour; cancer; tumour; treatment; expression; tumour suppressor;
 KW p53; pBR; ss.
 XX
 OS Pleurotus cornucopiae.
 XX
 PN JP11315096-A.
 XX
 PD 16-NOV-1999.
 XX
 PF 07-AUG-1998; 98JP-0236349.
 XX
 PR 08-AUG-1997; 97JP-0215311.
 PR 02-MAR-1998; 98JP-0066176.
 XX
 PA (NEWF-) NEW FOOD CREATION GIJUTSU KENKYU KUMIAI.
 XX
 WPI; 2000-058170/05.
 DR N-PSDB; AA246411.
 DR
 XX
 PT An antitumour protein derived from Pleurotus cornucopiae and its gene -
 useful for treatment of cancer including those caused by abnormal
 expression of cancer inhibitory gene (e.g. p53 and pBR)
 XX
 PS Claim 3; Page 13-15; 23pp; Japanese.
 XX
 CC The invention relates to a novel antitumour protein extracted from
 CC fruiting bodies of the fungus Pleurotus cornucopiae. The protein and
 CC nucleotides encoding it are useful for the treatment of cancer,
 CC including those caused by abnormal expression of tumour suppressor
 CC genes such as p53 and pBR. This sequence represents the antitumour
 CC protein.
 XX
 SQ Sequence 633 AA;

Query Match 39.2%; Score 1136; DB 21; Length 633;
 Best Local Similarity 45.3%; Pred.No. 1.8e-98;
 Matches 248; Conservative 70; Mismatches 175; Indels 54; Gaps 15;

QY 13 GYHKKEIEFQKIDREVNVIKALQOVSPVRNQNVPITLDPGAWSPGSSAISNGKNP 72
 Db 96 gshkntveyqknidkfvhlgqlmpvsvpnkvyadvdtlspasqa--sthfvrganp 153
 QY 73 HQREFENLSAEAVTRGVGGMSTHWCTPTRIHPHMESLPGIGRPKL-SNDPAEDDKWNE 131
 Db 154 eqdpfnlsggavtrvggmthwtcatprfhks-----erpklvkdddsadalewer 206
 QY 132 LYSEAEELICTSKEFDESIRHVLVLSLDAYKDRQIRFPLPLACHRLKNAPEYVEWH 191
 Db 207 lydlaesfvktghnqfdqsrhnlvleklqdesyag-grgfeqiplaaqrtn--prfvews 263

QY 192 SARNLF--HSIYNDKOKKLTLLTNHRCRLALTGGYERKKIGAAAEVRNLLATRNPSQL 249
 Db 264 sahtvfdlenrpnaddekgrfnlpavvcervt-rdsldrkiedievhdliag----- 315
 QY 250 DSY-INAQVTVVLASGAIGNPQILYNSGFSGLQVTPRND-----SLIPNLGRYITEQPMAPC 304
 Db 316 drykvkadvfilcsgavhnpqilvnsqf-grmgqpdsslppttllpylgsyiteqtlfc 374
 QY 305 QIVLROEFVDSVRDDP--YGLP-----WWKEAVAQOIAKKNPTDAL 342
 Db 375 qtvfstelvnlvksdmliivgtpgqdyvstftpdspenkhpnwnnekvqhmhmqdpl 434
 QY 343 PIPRDPPEQVTTPTTEHPHTQIHRDAFSYGAVGPEVDSRVIVDLRWFEGATDPEANNL 402
 Db 435 pipldpepqvttlfgdthpwtqihrdafsygavaesidrlvvdwrfgrtepvveen 494
 QY 403 LVFQNDVODGYSMFQPTPRYR--PSTASNVARKMADMCCEVASNIGGYLTPSPQFMDPG 461
 Db 495 lwfskqitdaynlpqtfisrfpggrtaqeaelmmadmtcmstkvggflpgsyppqfmapg 554
 QY 462 LALHLAGTTRIGFDKATTV--DNNSLVWDFANLYVAGNGTIRTGFCENPTLTSMCHAIK 519
 Db 555 lvhlgtghmgfdedkacvdtnskvmgnlfiggcgnigtayasnptltavala 614
 QY 520 SARSIIIN 526
 Db 615 sckyirn 621

RESULT 4
 AAR99628
 ID AAR99628 standard; Protein; 623 AA.
 XX
 AC AAR99628;
 XX
 DT 27-NOV-1996 (first entry)
 XX
 DE Pyranose oxidase.
 XX
 KW Pyranose oxidase; glucose; oxidation; glucanase; assay; diabetes;
 KW marker; diagnosis; 1,5-anhydro-D-sorbitol.
 XX
 OS Coriolus versicolor.
 XX
 PN DE19545780-A1.
 XX
 PD 13-JUN-1996.
 XX
 PF 07-DEC-1995; 95DE-1045780.
 XX
 PR 24-MAY-1995; 95JP-0124835.
 PR 07-DEC-1994; 94JP-0304086.
 XX
 PA (KIKK) KIKKOMAN CORP.
 XX
 PI Kawai G, Koyama Y, Minamihara T, Nishimura I, Okada K;
 PI Suzuki M;
 XX
 DR WPI; 1996-278990/29.
 DR N-PSDB; AAT34420.
 XX
 CC DNA encoding protein with pyranose oxidase activity at neutral pH -
 useful for the determination of glucose in body fluids or foods, or
 PT 1,5-anhydro-D-sorbitol used as marker for diabetes diagnosis
 XX
 PS Claim 1; Page 14-16; 22pp; German.
 CC The present sequence encodes a protein isolated from Coriolus versicolor,
 CC which has the enzyme activity of pyranose oxidase (PO). The PO oxidises
 CC glucose to glucosone and has an optimum pH of 7-7.5. It has a mol. wt. of
 CC 290000 (determined by gel filtration) and is stable at around 50deg.C.
 CC PO can be used for measurement of glucose in, e.g. foods or body fluids,

```
CC or 1,5-anhydro-D-sorbitol which is an important marker used in the
CC diagnosis of diabetes.
XX
SQ Sequence 623 AA;

Query Match 38.9%; Score 1128.5; DB 17; Length 623;
Best Local Similarity 45.0%; Pred. No. 8.8e-98;
Matches 250; Conservative 68; Mismatches 165; Indels 73; Gaps 15;

QY 13 GYHKKNEIEFKDIDRFVNVIKGALQOVSVPRNQNVTLPDGAWSAPPGSSAISNGKNP 72
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
87 gahkntveyqknidkfnvlgqqlmsvvpvntivldtspstswqa--ssffvrngsnp 144
QY 73 HQREFENLSAEAVTRGVGGMSTHWTCTSPRIHPPMESLPGIGRKLSDNPAE-DKKE 131
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
145 eqdplrnlgagvatrvvgmsthtcatprfdeq-----rpllvkddqdaadaewdr 197
QY 132 LYSEAEIRLIGSTKFEDESIRHTLVLSLODAYKDQRIFRPLPLACHRLKNAPYVEWH 191
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
198 lytkaesyfkttdqfkesirhnlvlnklaeeykg-qrdfgqiplaatr--rsptfvews 254
QY 192 SAENLF--HSIYNDKQKKLFTLLTNHRCRLALTGGEKKIGAEVRNLLATRNPSQL 249
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
255 santvfdlqnrpntdapnerfnlpavacery-----vrn-----tsnsei 295
QY 250 DSY-----IMARVYVLASGAIGNPQILYNSGFSGL-QVTPRN-DSLIPNLGRY 295
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
296 esalhndlisgrdfeikadvfvtgavhnaqlvnsfgqlgrpdpanpqlplsrsy 355
QY 296 TEQPMATCQIVLRQEFVDSVRDDPY--GLP-----WWKEAVAQAH 333
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
356 lteqalvfctvmsteliidvksdmilrgnpgdlqysvtytpgaetnkhpdwnekvknh 415
QY 334 TAKNETDALPIPRPEQVPTTPEEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRFG 393
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
416 mmqhgcdplpfepeqvvtlqfshpwhqihrdafsygavqgtidsrlivdrwifg 475
QY 394 ATDPKANLLVFQNDVQDGYSMQPQTFRRYR-PSTASNVRAKMMADMCEVASNLGGYLP 452
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
476 rtepeenklwfsdkitdntymppqtfdfirpagrtskeaedmmtcmvkwakigglfp 535
QY 453 SPQPMDCGLALHLAGTTRIGFDKA--TTVADNNSLVWDFANLYVAGNGTIRTFGENPT 510
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
536 alpqtmepglvhlhgtthrmgdeqecdcvntdsrvfgknlfggcgniptayganpt 595
QY 511 LTSMCHAIKARSIIIN 526
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
596 ltamslaiksczeyikn 611

RESULT 5
AAB48832
ID AAB48832 standard; Protein; 622 AA.
XX
AC AAB48832;
XX
DT 13-MAR-2001 (first entry)
XX
DE Trametes hirsuta pyranose oxidase.
XX
KW Pyranose oxidase; expression construct; recombinant production;
KW monosaccharide oxidation; 2-keto derivative;
KW hydrogen peroxide production.
XX
OS Trametes hirsuta.
XX
FN US6146865-A.
XX
PD 14-NOV-2000.
XX
PF 05-MAY-1999; 99US-0305381.
XX
```

```
PR 08-JUN-1998; 98DK-0000774.
PR 10-JUN-1998; 98US-0088724.
XX
PA (NOVO ) NOVO NORDISK AS.
XX
PI Schneider P, Christensen S, Lassen SF;
XX
DR WPI; 2001-049055/06.
DR N-PSDB; AAC87518, AAC87519.
XX
XX
PT Novel nucleic acid molecule encoding polypeptide having pyranose
PT oxidase activity used to design oligonucleotide probes to identify and
PT clone DNA encoding the polypeptide from different genera or species -
XX
PS Claim 5; Fig 1; 20pp; English.
XX
CC The invention relates to nucleic acids (e.g., AAC87518, AAC87519) which
CC encode Trametes hirsuta pyranose oxidase (AAB48832). The invention also
CC relates to expression constructs, expression vectors and recombinant
CC cells comprising pyranose oxidase nucleic acid sequences, and the
CC recombinant production of Trametes hirsuta pyranose oxidase. Pyranose
CC oxidase catalyses the oxidation of several monosaccharides in the
CC pyranose form at position C2 to produce 2-keto derivatives with the
CC release of hydrogen peroxide. Nucleic acids encoding Trametes hirsuta
CC pyranose oxidase may be used to produce the enzyme and to design
CC oligonucleotide probes to identify and clone genomic pyranose oxidase
CC cDNA or genomic DNA from different genera or species of microorganisms
CC (fungi or bacteria). The present sequence represents pyranose oxidase
CC from the fungus Trametes hirsuta.
XX
SQ Sequence 622 AA;

Query Match 38.6%; Score 1119; DB 22; Length 622;
Best Local Similarity 44.6%; Pred. No. 7e-97;
Matches 251; Conservative 69; Mismatches 183; Indels 60; Gaps 15;

QY 13 GYHKKNEIEFKDIDRFVNVIKGALQOVSVPRNQNVTLPDGAWSAPPGSSAISNGKNP 72
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
87 gshkntveyqknidkfnvlgqqlmpsvvpvntvtdtspaswqa--stffvrnganp 144
QY 73 HQREFENLSAEAVTRGVGGMSTHWTCTSPRIHPPMESLPGIGRKL-SNDPAEDDKWNE 131
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
145 eqdplrnlgagvatrvvgmsthtcatprfek-----lqrpilvknkskdaadewdr 197
QY 132 LYSEAEIRLIGSTKFEDESIRHTLVLSLODAYKDQRIFRPLPLACHRLKNAPYVEWH 191
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
198 lytkaesyfkttdqfkesirhnlvlnklaeeykg-vrdfgqiplaatr--qsptfvews 254
QY 192 SAENLF--HSIYNDKQKKLFTLLTNHRCRLALTGGEKKIGAEVRNLLATRNPSQL 249
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
255 sahtvfdlenrpnkdapqkrfnlfpavactnvrndnaseivg-ldvrdlbggksit--- 310
QY 250 DSYIMAKVYVLASGAIGNPQILYNSGFSGLQVTPRN-----SLIPNLGRYITEQPMAF 303
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
311 ---lkakvyilitagavhnaqlaasgfggl---grdpakplpsllpylgthiteqltfv 364
QY 304 CQIVLRQEFVDSVRDDP--YGLP-----WWKEAVAQAHIAKNPTDA 341
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
365 cqtvmstelsvntadmtivgkphdysvtvtytppgnpnknkhpdwnekvkxhmdhdqdp 424
QY 342 LPIPRDPEQVPTTPEEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRWFAGTDPKAN 401
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
425 lplfedpeqvvtlqfathpwhqihrdafsygavqgsidsrlivdrwifgrtepteen 484
QY 402 LLVFQNDVQDGYSMQPQTFRRYRPTASNVRAKMMADMCEVASNLGGYLPSPQFMDPG 461
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
485 klwfsdkitdaynlrqptfdfr--fpggreaedmmtcmvmsakigglfpgsyqfmeqpg 542
QY 462 LALHLAGTTRIGFDKAT--TVADNNSLVWDFANLYVAGNGTIRTFGENPTLTSCHAIK 519
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
543 lvlhlgthrmgfdkadkccvtdsdrvfgknlfggcgniptayaanptlamslaik 602
```

```
QY 520 SARSIIINTLKGGTGDKNYGEHRN 542
Db 603 sceyi-----kknfepnnpvkhn 622

RESULT 6
AAW87531
ID AAW87531 standard; Protein; 566 AA.
XX AC AAW87531;
XX DE 26-FEB-1999 (first entry)
XX DT An antitumour protein.
XX DE Antitumour protein; Tricholoma matsutake.
XX KW Tricholoma matsutake.
XX OS JP10313876-A.
XX PN 02-DEC-1998.
XX PD 13-FEB-1998; 98JP-0031452.
XX PF 13-FEB-1997; 97JP-0029275.
XX PR (MOMO-) MOMOYA KK.
XX PA (NORQ) NORINSUISANSO SHOKUHN SOGO.
XX XX WPI; 1999-074153/07.
XX DR N-PSDB; AAW83626.
XX XX
XX PT An anti-tumour protein - prepared by culture of host cell
XX PT transformed by vector containing base coding sequence
XX PS Claim 1; Page 7-8; 15pp; Japanese.
XX CC The present sequence encodes an antitumour protein, and is isolated
XX CC from Tricholoma matsutake.
XX XX
XX SQ Sequence 566 AA;

Query Match 35.3%; Score 1024.5; DB 20; Length 566;
Best Local Similarity 43.0%; Pred. No. 5.6e-88;
Matches 232; Conservative 77; Mismatches 180; Indels 51; Gaps 12;

QY 2 AEGTAVPVYGYHKKKEIEFQKIDRFVNVKIGALQOVVPVRNQVNPILDPGAWGAPP 61
Db 62 aeiysqdnvpvigaahhksnikfkdkfvnlingalqpsisdtvqptlavaaw-app 120
QY 62 GSSA-----ISNGKNPHOREFENLSAEAVTRGVGMSTHTWCSTPRIHPWESLPGIGRPK 117
Db 121 idpaegqlvngnhnpdeaglnlpgsavtrvcgvmachtwcactphd-----ee 170
QY 118 LSNDPAEDKEMNELYSEAEERLICTSTKEFDESIRHTLVLSLQADYKDRQIRPRLPLA 177
Db 171 rvnnpv-dkqefdalleraktllnvhsdqyddsirgqvketlqgtl-dasrgvttlplg 228
QY 178 CHRUKNAPEVWEHSAENLHSHIYNDQKQLFTLLNHRCTRALRGGEYKKTGAEEVR 237
Db 229 verrtdnpiytwgag-----tvlgdvpkspfralvtetrvtklivsetnptqvvaallr 284
QY 238 NLLATRPNSOLDYSYINAKVYVLASGAIGNPOILYNSGFSGLQVTPRNDLSLIPNLGRYIT 297
Db 285 nl-----ntsn--delvaksiviacgavctpqilwns-----nirpy-----algryls 327
QY 298 EQPMAFQIVLRQBEFVDSVRDDPYGLPWKEAVAQHIKNEPTDALPIPFRRDPEQVTTTF 357
Db 328 eqsmtfcqivlkgivdaiaatdpr----faakveahkkkbpddvlpipfhepeqvmipy 383
QY 358 TEEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRFWFGATDPANNLLVF-----QND 408
```

```
Db 384 tsdfpwhvqhrdafsdygdvgpkadprvvvdrlrfgksdiveenrvtfgnpklreweag 443
QY 409 VQDGYSMQPQTFERYRPSTASNVARKKMMADMCEVASNLGGVLTSPPOFMDPGLALHLAG 468
Db 444 vtdtygmppqptfkhvkrtnadgdrqmmndmtnvamlggylpqsyppqfnapglvlhitg 503
QY 469 TTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRTGEGNPTLTSMCHAKSARSIIINTL 528
Db 504 ttrigtddqtsvadptskvhfnfnlwgvgncipdatacnprrtsvayalkgaeavnyl 563

RESULT 7
AAB10457
ID AAB10457 standard; Protein; 564 AA.
XX AC AAB10457;
XX DE 11-DEC-2000 (first entry)
XX DT T. matsutake pyranose oxidase protein.
XX DE Pyranose oxidase; matsutake mushroom; antibacterial; sugar metabolism;
XX KW diabetes mellitus.
XX OS Tricholoma matsutake.
XX PN JP2000175698-A.
XX PD 27-JUN-2000.
XX PF 16-DEC-1998; 98JP-0357423.
XX PR 16-DEC-1998; 98JP-0357423.
XX PS (NISB) JAPAN TOBACCO INC.
XX DR WPI; 2000-478485/42.
XX DR N-PSDB; AAA71487.
XX PT A reagent containing pyranose oxidase for the determination of pyranose
XX PT
XX PS Claim 5; Page 17-18; 23pp; Japanese.
XX CC This invention describes a novel reagent containing pyranose oxidase for
XX CC the determination of pyranose which can be prepared from a fraction
XX CC precipitated from an aqueous extract of matsutake mushroom by ammonium
XX CC sulfate precipitation and has an antibacterial activity against at least
XX CC Pyricularia oryzae or Thanatephorus cucumeris and has a M.W. of ca.
XX CC 210 kD by gel filtration and shows the presence of components of ca.
XX CC 50 kD and ca. 15 kD by SDS-PAGE and maintains the above antibacterial
XX CC activity by being heated at 60 degrees C for 10 minutes in a neutral
XX CC aqueous solution and in which the above antibacterial activity is
XX CC inactivated by being heated at 80 degrees C for 10 minutes in a neutral
XX CC aqueous solution. The invention also describes a method for the
XX CC determination of pyranose in a sample in which the above reagent for the
XX CC determination of pyranose is reacted with pyranose in the sample and the
XX CC hydrogen peroxide formed is reacted with an enzyme to develop a color, a
XX CC method for diagnosing a disease accompanied by abnormality in sugar
XX CC metabolism in which the above reagent for the determination of pyranose
XX CC is reacted with pyranose in the sample and the hydrogen peroxide formed
XX CC is reacted with an enzyme to develop a color, and a kit for pyranose
XX CC analysis or the diagnosis of a disease accompanied by abnormality in
XX CC sugar metabolism. The pyranose oxidase can be used as a diagnostic agent
XX CC for diabetes mellitus. This sequence represents the Tricholoma matsutake
XX CC pyranose oxidase protein which is described in the method of the
XX CC invention.
XX SQ Sequence 564 AA;
```

Query Match

33.7%; Score 976.5; DB 21; Length 564;

Best Local Similarity 41.3%; Pred. No. 2e-83;
Matches 223; Conservative 82; Mismatches 182; Indels 53; Gaps 13;

QY 2 AEEGTAVPVVPGVHKNETEFKIDRFVNVKLGALQOVSVPRNQNVPTLDPGAWSAAPP 61
DB 62 aelsgqdnvlgahhrnsikfgtdkdfvniingalqpsispsdyqptlavaaw-app 120
QY 62 GSSA-----ISGNKPNQREFENLSAEAVTRGVGGMSTHTWCTSPRIHPMESLPGIGRPK 117
DB 121 idpaegqlvimhnpqeaqlnlpgsavtrtvgmawhwtacphtd-----ee 170
QY 118 LSNDAEDDKENNELYSAEERLIGTSTKEFDESIRHTLVLSLDQAYKDRQRIPLPLA 177
DB 171 rvnnpv-dkqefdalleraktllnvhsdgydsirgivvketlqtl-dasrgvttlplg 228
QY 178 CHRLKNAPEYVWHSAENLFHSIYNDKQKLFLLTNHRCRLALTGGYEKKIGAAEVR 237
DB 229 verrtdnplyvtwtgad----tvlgdvkpsrfrvltvtrtkflvsetnptqvvaaalr 284
QY 238 NLLATRNPSQLDSYIMAKVYVYVLAIGAIGNPQILYNSGFSGLQVTPRNDSLIPNLGRYIT 297
DB 285 nl-----ntsn--delvvaqsviaagavctpqilwns-----nirph-----algryls 327
QY 298 EQPMAPFCQILVLRQEFVDSVRDDPYGLPWKKEAVAQHIKKNPTDALPIPRDPPEQVTPPF 357
DB 328 eqsmtfcqilvkrsvdsiatdpr----faakveahkkkhpddvipiphepepvmipy 383
QY 358 TEEHPWHTQIHRDAFYSYGAAGPEVDSRVIVDLRWFSGATDPEANNLLVF-----QND 408
DB 384 tsdfpwhvqvr--yafgdvpgkadrprvvvdlrffgksdiveenrvtfpgnpklrdweag 441
QY 409 VODGYSMPQPTFRYRSTASNVRAKMMADMCCEVASNLGGYLTSPQPMDFGLALHLAG 468
DB 442 vldtygmpeptfhvkrtnadgrdrmmndmtvnaniigylpgsyppqfmapaglaqhtg 501
QY 469 TTRIGFDKATTVADNNSLVWDFANLYVAGNGVIRTFGFGENPFLTSMCHAIKARSIIINTL 528
DB 502 ttrigtddqtsvadptskvhnfdnlwvgngcipcdatacnprtstsvayalkgaeavvsyl 561

RESULT 8
AAW81952
ID AAW81952 standard; Protein; 564 AA.
XX
AC AAW81952;
XX
DT 07-JUL-2000 (first entry)
XX
DE Trichoderma derived antifungal protein sequence.
XX
KW Antifungal protein; Pyricularia oryzae; Rhizoctonia solani; mushroom;
KW growth inhibitor; plant pathogenic fungi; antibacterial agent;
KW N-terminal fragment.
XX
OS Trichoderma matsutake.
XX
SN WO200014242-A1.
XX
PD 16-MAR-2000.
XX
PF 19-AUG-1999; 99WO-JP04441.
XX
PR 08-SEP-1998; 98JP-0270606.
XX
PA (NTSB) JAPAN TORACCO INC.
PA (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
PI Takakura Y, Kuwata S, Ohta S;
XX
DR WPI; 2000-256990/22.
DR N-PSDB; AAA07403.
XX
PT Mushroom-derived antibacterial protein against plant pathogenic fungi

PT of rice, with activity and thermal stability, obtainable cheaply on
XX large scale, useful in agriculture -
PS Claim 3; Page 41-45; 52pp; Japanese.
XX
CC This sequence is the Trichoderma matsutake antibacterial protein of
CC the invention. The protein has activity against at least Pyricularia
CC oryzae and Rhizoctonia solani, and is obtained from a fraction of an
CC aqueous extract of a mushroom precipitated by the ammonium sulphate
CC precipitation method. The protein has a molecular weight of about 210 kD
CC as determined by the gel filtration method, includes components of about
CC 15 kD and 50 kD in SDS-PAGE, and is stable to heating in an aqueous
CC neutral solution at 60 degrees C for 10 minutes but with loss of
CC antibacterial activity after heating in the solution at 80 degrees C for
CC 10 minutes. The protein is used for inhibiting the growth of plant
CC pathogenic fungi e.g. Pyricularia oryzae and Rhizoctonia solani.
CC It is useful in treating rice plants, and is applicable in agriculture as
CC an antibacterial agent. The protein has activity at relatively low
CC concentrations, and can be produced at low cost on large scale.
XX
SQ Sequence 564 AA;

Query Match 33.7%; Score 976.5; DB 21; Length 564;
Best Local Similarity 41.3%; Pred. No. 2e-83;
Matches 223; Conservative 82; Mismatches 182; Indels 53; Gaps 13;

QY 2 AEEGTAVPVVPGVHKNETEFKIDRFVNVKLGALQOVSVPRNQNVPTLDPGAWSAAPP 61
DB 62 aelsgqdnvlgahhrnsikfgtdkdfvniingalqpsispsdyqptlavaaw-app 120
QY 62 GSSA-----ISGNKPNQREFENLSAEAVTRGVGGMSTHTWCTSPRIHPMESLPGIGRPK 117
DB 121 idpaegqlvimhnpqeaqlnlpgsavtrtvgmawhwtacphtd-----ee 170
QY 118 LSNDAEDDKENNELYSAEERLIGTSTKEFDESIRHTLVLSLDQAYKDRQRIPLPLA 177
DB 171 rvnnpv-dkqefdalleraktllnvhsdgydsirgivvketlqtl-dasrgvttlplg 228
QY 178 CHRLKNAPEYVWHSAENLFHSIYNDKQKLFLLTNHRCRLALTGGYEKKIGAAEVR 237
DB 229 verrtdnplyvtwtgad----tvlgdvkpsrfrvltvtrtkflvsetnptqvvaaalr 284
QY 238 NLLATRNPSQLDSYIMAKVYVYVLAIGAIGNPQILYNSGFSGLQVTPRNDSLIPNLGRYIT 297
DB 285 nl-----ntsn--delvvaqsviaagavctpqilwns-----nirph-----algryls 327
QY 298 EQPMAPFCQILVLRQEFVDSVRDDPYGLPWKKEAVAQHIKKNPTDALPIPRDPPEQVTPPF 357
DB 328 eqsmtfcqilvkrsvdsiatdpr----faakveahkkkhpddvipiphepepvmipy 383
QY 358 TEEHPWHTQIHRDAFYSYGAAGPEVDSRVIVDLRWFSGATDPEANNLLVF-----QND 408
DB 384 tsdfpwhvqvr--yafgdvpgkadrprvvvdlrffgksdiveenrvtfpgnpklrdweag 441
QY 409 VODGYSMPQPTFRYRSTASNVRAKMMADMCCEVASNLGGYLTSPQPMDFGLALHLAG 468
DB 442 vldtygmpeptfhvkrtnadgrdrmmndmtvnaniigylpgsyppqfmapaglaqhtg 501
QY 469 TTRIGFDKATTVADNNSLVWDFANLYVAGNGVIRTFGFGENPFLTSMCHAIKARSIIINTL 528
DB 502 ttrigtddqtsvadptskvhnfdnlwvgngcipcdatacnprtstsvayalkgaeavvsyl 561

RESULT 9
AAW87541
ID AAW87541 standard; Peptide; 51 AA.
XX
AC AAW87541;
XX
DT 26-FEB-1999 (first entry)
XX
DE Peptide derived from an antitumour protein.

PS Claim 2; Page 27; 52pp; Japanese.

XX This sequence represents an N-terminal fragment of the protein of the invention. The protein is an antibacterial protein with activity against at least *Pyricularia oryzae* and *Rhizoctonia solani* is obtained from a fraction of an aqueous extract of a mushroom precipitated by the ammonium sulphate precipitation method. The protein has a molecular weight of about 210 kD as determined by the gel filtration method, includes components of about 15 kD and 50 kD in SDS-PAGE, and is stable to heating in an aqueous neutral solution at 60 degrees C for 10 minutes but with loss of antibacterial activity after heating in the solution at 80 degrees C for 10 minutes. The protein is used for inhibiting the growth of plant pathogenic fungi e.g. *Pyricularia oryzae* and *Rhizoctonia solani*. It is useful in treating rice plants, and is applicable in agriculture as an antibacterial agent. The protein has activity at relatively low concentrations, and can be produced at low cost on large scale.

XX Sequence 67 AA;

Query Match 6.1%; Score 177; DB 21; Length 67;
 Best Local Similarity 57.4%; Pred. NO. 4.4e-09;
 Matches 35; Conservative 6; Mismatches 20; Indels 0; Gaps 0;

QY 409 VODGYSMQPTRYRSTASNVRAKMKMADCEVASNLGGVILPTSPQPMDFGLALHLGAG 468
 Db 7 vcdtympqptfhvkrtnadgdrdmndmtvnmggylpqsyypqfmapglvlhitg 66

QY 469 T 469
 Db 67 t 67

RESULT 12
 AAB97034
 ID AAB97034 standard; Peptide; 34 AA.
 AC AAB97034;
 XX
 DT 20-JUL-2001 (first entry)
 XX
 DE Lyophyllum shimeji antibacterial protein fragment #4.
 XX
 KW Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;
 KW *Pyricularia oryzae*; *Rhizoctonia solani*; rice pathogen.
 XX
 OS Lyophyllum shimeji.
 XX
 PN WO200121657-A1.
 XX
 PD 29-MAR-2001.
 XX
 PF 20-SEP-2000; 2000WO-JP06404.
 XX
 PR 21-SEP-1999; 99JP-0267238.
 XX
 PA (NTSB) JAPAN TOBACCO INC.
 PA (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
 PI Takakura Y, Kuwata S, Inoue Y;
 XX
 DR WPI: 2001-281598/29.
 DR N-FSDB; AAF99980.

Antibacterial protein and encoded gene isolated from *Lyophyllum shimeji*, with activity against plant pathogenic bacteria, applicable in agriculture e.g. rice cultivation at low concentration, produced at low cost on large scale -

Claim 2; Page 24; 52pp; Japanese.

The present sequence is part of an antibacterial protein from the fungus *Lyophyllum shimeji*. The protein was obtained from a fraction prepared by extracting *Lyophyllum shimeji* with water and subjecting the extract to ammonium sulphate precipitation. The protein inhibits the growth of the plant pathogenic bacteria *Pyricularia oryzae* and *Rhizoctonia solani* at a relatively low concentration. P. oryzae and R. solani are causative of the two major diseases of rice. The protein contains components of 70 kDa and 65 kDa, as determined by SDS-PAGE. The antibacterial protein can be produced at low cost on a large scale.

Query Match 5.6%; Score 162; DB 22; Length 30;
 Best Local Similarity 100.0%; Pred. No. 3e-08;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX Sequence 30 AA;

CC to ammonium sulphate precipitation. The protein inhibits the growth of the plant pathogenic bacteria *Pyricularia oryzae* and *Rhizoctonia solani* at a relatively low concentration. P. oryzae and R. solani are causative of the two major diseases of rice. The protein contains components of 70 kDa and 65 kDa, as determined by SDS-PAGE. The antibacterial protein can be produced at low cost on a large scale.

XX Sequence 34 AA;

Query Match 5.8%; Score 168; DB 22; Length 34;
 Best Local Similarity 100.0%; Pred. No. 1e-08;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 AERLIGTSTKEFDESIRHTLVRLSLQDAYKDRQR 169
 Db 1 aerlgtstkefdesirhtlvrlslqdaykdrqr 34

RESULT 13
 AAB97031
 ID AAB97031 standard; Peptide; 30 AA.
 XX
 AC AAB97031;
 XX
 DT 20-JUL-2001 (first entry)
 XX
 DE Lyophyllum shimeji antibacterial protein fragment #1.
 XX
 KW Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;
 KW *Pyricularia oryzae*; *Rhizoctonia solani*; rice pathogen.
 XX
 OS Lyophyllum shimeji.
 XX
 PN WO200121657-A1.
 XX
 PD 29-MAR-2001.
 XX
 PF 20-SEP-2000; 2000WO-JP06404.
 XX
 PR 21-SEP-1999; 99JP-0267238.
 XX
 PA (NTSB) JAPAN TOBACCO INC.
 PA (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
 PI Takakura Y, Kuwata S, Inoue Y;
 XX
 DR WPI: 2001-281598/29.
 DR N-FSDB; AAF99980.

Antibacterial protein and encoded gene isolated from *Lyophyllum shimeji*, with activity against plant pathogenic bacteria, applicable in agriculture e.g. rice cultivation at low concentration, produced at low cost on large scale -

Claim 2; Page 32; 52pp; Japanese.

The present sequence is part of an antibacterial protein from the fungus *Lyophyllum shimeji*. The protein was obtained from a fraction prepared by extracting *Lyophyllum shimeji* with water and subjecting the extract to ammonium sulphate precipitation. The protein inhibits the growth of the plant pathogenic bacteria *Pyricularia oryzae* and *Rhizoctonia solani* at a relatively low concentration. P. oryzae and R. solani are causative of the two major diseases of rice. The protein contains components of 70 kDa and 65 kDa, as determined by SDS-PAGE. The antibacterial protein can be produced at low cost on a large scale.

Query Match 5.6%; Score 162; DB 22; Length 30;
 Best Local Similarity 100.0%; Pred. No. 3e-08;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX Sequence 30 AA;

Query Match 5.6%; Score 162; DB 22; Length 30;
 Best Local Similarity 100.0%; Pred. No. 3e-08;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 15:51:06 ; Search time 85.12 Seconds
(without alignments)
1256.002 Million cell updates/sec

Title: US-09-856-327-2
Perfect score: 3284
Sequence: 1 MSLSSTQMLRDPKSMQING.....INTLKGTDGKNTGHRNL 618

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vtebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1232.5	37.5	623	3 P79076	P79076 coriolus ve
2	203	6.2	551	2 Q9RH54	Q9RH54 pantoea agg
3	200.5	6.1	573	16 Q9PI90	Q9PI90 campylobact
4	200	6.1	545	2 Q9LBK8	Q9LBK8 gluconobact
5	191.5	5.8	615	2 O34214	O34214 peccobacter
6	190.5	5.8	579	16 Q9A7T6	Q9A7T6 caulobacter
7	182	5.5	591	16 Q9I1K8	Q9I1K8 pseudomonas
8	180	5.5	553	2 Q9XCR0	Q9XCR0 pantoea cit
9	179.5	5.5	529	17 Q9HQH8	Q9HQH8 halobacteri
10	176	5.4	722	16 Q9RZ26	Q9RZ26 deinococcus
11	158.5	4.8	748	10 Q9M0H4	Q9M0H4 arabidopsis
12	158.5	4.8	748	10 Q9ABP3	Q9ABP3 arabidopsis
13	155.5	4.7	499	16 Q988P2	Q988P2 rhizobium 1
14	155.5	4.7	769	3 O74253	O74253 pycnopus
15	149	4.5	479	16 Q98D66	Q98D66 rhizobium 1
16	141	4.3	523	16 Q98C76	Q98C76 rhizobium 1

17	138.5	4.2	768	3 O42729	O42729 trametes ve
18	127.5	3.9	578	16 Q57307	Q57307 mycobacteri
19	126.5	3.9	772	3 Q12661	Q12661 phanerocoe
20	124.5	3.8	578	16 Q92XY0	Q92XY0 rhizobium m
21	123	3.7	531	16 Q916D0	Q916D0 pseudomonas
22	123	3.7	704	3 Q9P8D9	Q9P8D9 candida tro
23	119	3.6	746	10 Q9LW56	Q9LW56 arabidopsis
24	118.5	3.6	785	3 Q9P8H5	Q9P8H5 humicola in
25	117.5	3.6	620	3 Q12623	Q12623 humicola gr
26	117	3.6	736	10 Q9ZWB9	Q9ZWB9 arabidopsis
27	116.5	3.5	698	3 Q9P8D8	Q9P8D8 candida clo
28	116	3.5	502	5 O18672	O18672 caenorhabdi
29	116	3.5	599	5 Q18429	Q18429 caenorhabdi
30	115.5	3.5	502	16 Q92L01	Q92L01 rhizobium m
31	114	3.5	1012	10 O64492	O64492 arabidopsis
32	113.5	3.5	678	10 O65709	O65709 arabidopsis
33	112	3.4	520	16 Q988D6	Q988D6 rhizobium l
34	112	3.4	828	3 O74240	O74240 thielavia h
35	112	3.4	1011	2 Q9L5M2	Q9L5M2 salmonella
36	112	3.4	1011	2 Q935H0	Q935H0 salmonella
37	112	3.4	1881	12 Q9DUN3	Q9DUN3 vesicular e
38	111.5	3.4	1899	5 Q9NDY7	Q9NDY7 leishmania
39	110.5	3.4	507	16 Q982L9	Q982L9 rhizobium l
40	110.5	3.4	589	10 Q9M4V5	Q9M4V5 oryza sativ
41	110.5	3.4	5170	5 Q17490	Q17490 caenorhabdi
42	110.5	3.4	6994	5 Q17343	Q17343 caenorhabdi
43	109.5	3.3	1125	5 Q93203	Q93203 caenorhabdi
44	109	3.3	2193	12 Q9QF31	Q9QF31 human coxa
45	108	3.3	1498	16 Q92FP9	Q92FP9 listeria in

ALIGNMENTS

RESULT	1
P79076	
ID	P79076
AC	P79076; PRELIMINARY; PRT; 623 AA.
DT	01-MAY-1997 (TrEMBLrel. 03, Created)
DT	01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	PYRANOSE OXIDASE.
OS	Coriolus versicolor.
OC	Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC	Aphyllophorales; Corioliaceae; Coriolus.
OX	NCBI_TaxID=57466;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=PS4A;
RX	MEDLINE=971177816; PubMed=9025322;
RA	Nishimura I., Okada K., Koyama Y.;
RT	"Cloning and expression of pyranose oxidase cDNA from Coriolus
RT	versicolor in Escherichia coli.";
RL	J. Biotechnol. 52:11-20(1996).
DR	EMBL; D/3369; BAA1119.1; -
DR	InterPro; IPR00205; NAD_binding.
SQ	SEQUENCE 623 AA; 69495 MW; 5D3FC81B35FA5B54 CRC64;

Query Match	37.5%;	Score	1232.5;	DB 3;	Length	623;			
Best Local Similarity	45.0%;	Pred. No.	2.7e-91;						
Matches	273;	Conservative	73;	Mismatches	176;	Indels	85;	Gaps	16;
QY	37	DVFTAGSGIGATVAKLCV	EAGLRVNVNVEIGADSFYAVNAEECTAVPYVPVGYHKKNEIE	96					
Db	48	DVIVGSGPIGCTYARELV	EAGYKVMFDEIDSGLKI-----GAHKKNIVE	95					
QY	97	FKQDIDRFVNVIKALQOV	SVPNQVPTLDPGAWSPGSSAISNGKNPHQREFNLS	156					
Db	96	YQKNIDRFVNVIOGLMSV	PNVTLVIDLSPISWQA--SSFFVRNGSNPEQDPLNLS	153					
QY	157	AEAVTRVGGMSTHWTCT	STRIHPHPPMESLPGIGRPKLSNDPAE--DKKEWELNLYSEARLI	215					

```

Db 154 GQAVTRVVGGMSTHTWCATPFRDRQ-----RPLLVKDDQDADDAEDWRLTYTKAESYF 206
Qy 216 GTSTKEFDESIRHTLVLRSLDAYKDRQIRFRPLPLACHRLKNAPEYVEWHSAENLF--H 273
Db 207 KTGTDOFKESIRHNLVLNKLAEYKG-QRDFQIPLAATR--RSPTFEVWSSANTVFDLQ 263
Qy 274 SIYNDKOKKFLTTLTNHRCIRLALTGGYEKKIGAAEVRNLLATRNPSQLDSY----- 327
Db 264 NRPNTDAPNERFNLFPAVACERV-----VRN-----TSNSEIESLHIDLI 304
Qy 328 -----INAKVVVLASGAIGNPOILYNSGESGL-QVTPRN-DSLIPNLGLRYITEQPMAF 379
Db 305 SGDRFEIRADVFVLTAGAVHQAQLLVNSGFGOLGRDPDPANPQLLPSSLGSIYTEQSLVFC 364
Qy 380 QIVLRQEFVDSVRDDPY--GLP-----WKEAVAQHIKAKNPTDAL 417
Db 365 QTVMSTELIDSVKSDMIIRGNPGDLGYSVTYTPGAETNKHDPWNNEKYKNHMMQHOEDPL 424
Qy 418 PIPFRDPPEQVTTPTTEBHPNHTQIHRDAFSYGAVGPEVDSRVIVDLRWFGATDPENNL 477
Db 425 PIPFEDPEQVTTTLFPQSPHPWHTQIHRDAFSYGAVQQSIDSRLLIVDWRFFRGTEPEKENK 484
Qy 478 LVFONDVDGYSMPQPTFRYR- PSTASNVARKMMADMCVASNLGGYLPTSPPOFMDPG 536
Db 495 LWFSDKITDTYNNMPQPTDFRPAGRTSKEADMMTMCVMSAKTGGFLPGSLPQFMPEPG 544
Qy 537 LALHLAGTRTGFQDKA--TTVADNNSLVMDFANLYVAGNGTIRTFGFGNPNTLSCHAIK 594
Db 545 LVHLGGTHRMGFDEQDKCCVNTDSRVFGFNFLGCGGNIPRAYGANPILATMSLAIK 604
Qy 595 SARSIIN 601
Db 605 SCEVIKN 611

RESULT 2
Q9RH54
AC ID Q9RH54 PRELIMINARY; PRT; 551 AA.
CD 09RH54;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE DEHYDROGENASE SUBUNIT I,
OS Pantoea agglomerans.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RA Chang K.-I., Yum D.-Y., Pan J.-G., Shin Y.-C.;
RT "Cloning and expression of a gene cluster encoding three subunits of
RT membrane-bound 2-keto-D-gluconate dehydrogenase from Erwinia herbicola
RT ATCC08111 in Escherichia coli.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF068066; AAF21261.1; -.
DR InterPro; IPR000205; NAD_binding.
SQ SEQUENCE 551 AA; 60047 MW; FE2ECCD0933DAFBF CRC64;

Query Match 6.2%; Score 203; DB 2; Length 551;
Best Local Similarity 19.9%; Pred. No. 8.7e-08;
Matches 132; Conservative 84; Mismatches 250; Indels 198; Gaps 26;

Qy 24 KNATHEYTNGDGVDFIAGSGPIGATYAKLCVEAGLRVWVVEIGANDSFYAVNAEETAV 83
Db 2 KKPVFTAGQASADIVVIGSGIVGMMANELVSGQSYLVLEAGL----- 46
Qy 84 PYVPCYHKKEIEFQKIDRVNVNVIKALQOVSVPRNQ-----NVPTLDPGAWSAP----- 135
Db 47 -----RIDRAQAVENRN-----HPFANRAGSDQGLLPQSKFAPAPLYF 86
Qy 136 PGSAISNGKNPHQREFENLSAEAVTRGVGMSTHTWCATPFRDRQ-----MESLPGIGRPK 192

```

Db	87	PRNNYV-NVTGPNADSFQ----	QCYLRTVGGTTHHWAASCRHHHPDSFVMSQSYGYGVR--	133
Qy	193	LSNDPAEDKREW-----NEL----	YSEARLIGTSTKEFDESIRHTLVLRSLQDAYKDRQRI	245
Db	140	-----DWPIGYDELEPWYCKAEINIGVAGP--NDPARQSPTRS-----	176	
Qy	246	FRPLPLACHRLKNAPEYVEHWSAENLFHSYINDOKKKL-----FTLLLTNH	291	
Db	177	-QPYPM-----DMVPPFAHGDNYFASVYNPHGYNLVP	226	
Qy	292	RCTRLALTGGVEKKIGAAEYRNLLATRNPSQLDSYIM-----	329	
Db	227	NCQPCIPGAMYNGIHVFE-----RAERNGAVVLAEEVYKMDTSDNNRITAVHWLDTSGA	282	
Qy	330	-----AKVYVILASGAIGNPOILYNSGFSGLQVTPRNDSLIPN----	380	
Db	283	SHKATAKAFALACNGIETPRLL-----MAANDANPNGIANASDMVGRNMDHSGFHCS	336	
Qy	381	IVLRQEFVDSVRDPPYGLPWKKEAQAQHIAKNTP-DALPIPRDRPPEQVTPPTTEHPWH	439	
Db	337	FLTKR-----PVM-----LGKGAQSCMWGYRGD-----FRDYSAN	370	
Qy	440	TQI-----HRDAFSYGAVGPEYDSRV-----IVDLRFWFGATDPEANLLVFQND	483	
Db	371	KVILNNISRVVTTAQAMKGLGVKALDEEIRYRAVHSVDLSISLEPLDPDENRLFLSKT	430	
Qy	484	VDCGYSPMQPTFRYPSTASNRARKMADCEVASNLGYLTPSPQFMDPGLAL--HL	541	
Db	431	RKDPHGLPCPDYYVDYGVYRKGAESAHAQLEHI-----GQLFDAKEFTISQGLNANNHI	485	
Qy	542	AGTTTRIGFDKATTVADNNSLVDFANLYVAGNCTIRGTGCGENTLFSMCHAIKSARSII	601	
Db	486	MGVIMGKNKAKEAVVDGNCRAFDHENLWLPGGGAIPASVYVNSTLTMAALGLKAAHDISL	545	
Qy	602	TLKG 605		
Db	546	RMKG 549		
RESULT	3			
Q9PI90				
ID	Q9PI90	PRELIMINARY;	PRT;	573 AA.
AC	Q9PI90;			
DT	01-OCT-2000 (T-EMBLrel. 15, Created)			
DT	01-OCT-2000 (T-EMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (T-EMBLrel. 19, Last annotation update)			
DE	PUTATIVE OXIDOREDUCTASE SUBUNIT.			
GN	CJ0415.			
OS	Campylobacter jejuni.			
OC	Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;			
OC	Campylobacter.			
OX	NCBI_TaxID=197;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=NCCT 11168;			
RX	MEDLINE=20150912; PubMed=10688204;			
RA	Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,			
RA	Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtby S.,			
RA	Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,			
RA	Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,			
RA	Whitehead S., Barrett B.G.;			
RT	"The genome sequence of the food-borne pathogen Campylobacter jejuni			
RT	reveals hypervariable sequences."			
RL	Nature 403:665-668(2000).			
DR	EMBL; ALI39075; CAB74251.1;			
DR	Complete proteome.			
SW	SEQUENCE 573 AA; 63686 MW; 7BEB522A779A5F94			
SEQ	CRC64;			

Query Match	6.1%	Score 200.5;	DB 16;	Length 573;
Best Local Similarity	20.6%	Pred. No. 1.5e-07;		
Matches 135; Conservative	74;	Mismatches 270;	Indels 175;	Gaps 26;


```
FT CHAIN 23 615 DEHYDROGENASE SUBUNIT.
SQ SEQUENCE 615 AA; 67241 MW; B9E1A84FD035609A CRC64;

Query Match
Best Local Similarity 5.8%; Score 191.5; DB 2; Length 615;
Matches 134; Conservative 83; Mismatches 256; Indels 163; Gaps 32;

QY 36 VDVFIAGSPIGATYAKLCVEAGLRVWVEIGA-----ADSFYAVNABEGTAVPVVPGVH 90
   ||| : : : ||| ||| ||| ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 29 VDAVVVGFAGAGAIMAKELTEAGLNVALERGPHRDTYPDGAYPOSIDELT-----YN 81

QY 91 KNEIEFQKIDRFNVNVIKGAQQVSVVRNQNVTPLDPAWSAPPGSSAISNGKNPHQR 150
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 82 IRKKL-FQ-DLSKSVTTRHDSQATVYR-----109

QY 151 EFNLSAEAVTGVGSMSTHTCSTPRIHPMESLPG-----IGRPKLSNDPAEDD--KE 203
   ||| : : : ||| ||| ||| ||| ||| : : : ||| : : : ||| : : : |||
Db 110 ---QLAAFLPGTGTGAGLHNSGVHFRVDPVELNLSRHYEARVKNFIPEGMTIQDFGVS 166

QY 204 WNEL---YSEARLIGTSTKEFDESRHTLVLSRLQDAY--KDRQRIFRPLP-----LAC 253
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 167 YNELEPFQAEKVFGTSSAW--TIKMGKGEKGNFYAPDRSSDF-PLPAQKRTYSA 223

QY 254 HRLKNAPEYVEWH-----SAENLFHSIYND-----279
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 224 QLFAQAABSVGYHPYDMPESANTSGPYNTYGAQMPCNFCGYCSGYACYMYSKASPNVNI 283

QY 280 ----KOKLFTLLTHRCTRLALTGGEKKIGA-----AEVRNLLATRNPSQ--LDSYM 329
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 284 LPALROEPKFEURNAYLVRLNLTGDKKRATGVYLDGQGEV---QPADIVILSAFQF 340

QY 330 AKVYVLASGAIGNP-QILYNSFGSLQVTPRNDLSIPNL-GRYITEQPM--AFQOIVLR 384
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 341 HNVHMLLSGIGQPNITNEGVGRNAYQNIISLTKALFDKNTTNFEGAGGAGVAVD 400

QY 385 QEFVSDVRDDPYGL-----PWWKEAVAQIAKNPTDALPIPRDPPEQVTPPTTEHPMHT 440
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 401 DFNADNFHDPGYGFVGGSPFW---VNOAGTKPVSGLPKTPN-----WGS 444

QY 441 QIHRDAFSGVAGPEVDSRVIVDLRFWFGATDPEANLLVFQNDVODGYSP--QPTFRYR 498
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 445 QWK-----AAVDYTNHHSIDAH--GAHOSYRANYLDLDPNYKNVYQPLLRMTFDMQ 496

QY 499 PSTASNVR-ARKMADMEV--ASN-----LGGYLTPSPQFMDPGL--ALHLAGTTRIGF 549
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 497 D---NDIRMAQPMVKMRKITEAMPKMLIG--AKGPTGHTDTTYQYQTHMSGGALIGE 551

QY 550 DKATTVDNNSLVDFANLIYVAGNTIRTGFGENTP 585
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 552 DPKTSAVNRYLQSDVDPNVFVPGASAFPQGLGYNPT 587

RESULT 6
QY9A7T6 PRELIMINARY; PRT; 579 AA.
AC Q9A7T6;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE OXIDOREDUCTASE, GMC FAMILY.
GN CC1634.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Pollock I., Nelson W.C., Newton A., Stephens C., Madhke N.D., Ely B.,
```

```
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005839; AAK23612.1; -.
DR TIGR; CC1634; -.
KW InterPro; IPR000205; NAD_binding.
KW Complete proteome.
SQ SEQUENCE 579 AA; 64397 MW; 7125C4DFAD618F10 CRC64;

Query Match
Best Local Similarity 5.8%; Score 190.5; DB 16; Length 579;
Matches 136; Conservative 83; Mismatches 270; Indels 185; Gaps 28;

QY 18 INQTPKNAIETHYNDGVDVFIAGSPIGATYAKLCVEAGLRVWVEIGAADSFYAVNA 77
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 4 LNGRARRK---NTY-----DAIVVSGITGGIAAKELTEKGLKLVL-----42

QY 78 EBGTAVPVPGVYHKKNEIEFQKIDRFNVNVIKGAQQVSVVRNQNVT--LDPCAWSAP 135
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 43 ERGPMVRHLEDY-----PTAMLDPMQSKYP 67

QY 136 PGSSA-----ISNGKNPHQR--FENLSAEAVTRG--VG 165
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 68 QGKLEAEALNAHYKVQRTGYTWTQTQHFVVRDDEHYTEENRFDWI-----RGYHVG 121

QY 166 GMSHTWTCSTPRIHPMESLPGIGRPKLSND---PAEDDKENNELYSEARLIGTSTKEF 222
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 122 GRSLTWGRQSVR-HSPID-FEANAEGIAVDWPIRYDLAPW---YEHVERFIGVSGQA- 175

QY 223 DESIRH-----TLVLSRLQDAYKDRORI-FRPLPLACHRLKNAPEVEVHSA----- 268
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 176 -EGLPHFPDGHVQPPMELNCVEKAFKARSEARFPERRVTIGRTAHLTDPTEDQLALGRTK 234

QY 269 ---ENL-----FHSIYNDKOKKLTLLTNHCTRLALTGG-----YEKKIGAAEVR 312
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 235 COYRNLICIRGCPFGAYSSNSG---GLIAERTGNLVRPNSIYVELIYDERAGRASCV 290

QY 313 NLLATRNPSQLDSYIMAKVYVVLASGATGNPOILYNSFGSLQVTPRNDLSIPN----- 366
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 291 RIL---DAETRDEEFHADVIFLCASALNSAWIMNS-----TSSRFPNFGNAS 337

QY 367 --LGRYITEQPMACQIVLRQEFVDSV---RDDPYGLPWKEAQAQHTAKNPTDALPIP 420
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 338 DQLGRNVMDHHLGAGATGAQPEFADMYFSGRPNIGYVPRFRN-LGDAASKRSDYLRFSG 396

QY 421 FRDPEQVTPPTTEHPMHTQIHRDAFSYGAV-----GPEVDSRVIVDLRFWFGATDP 472
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 397 YQGAGRAT-----WERDRGQGGRGFGAARKAALSQGPWT-----MGLSGFGEMLP 443

QY 473 EANNLLVFQNDVODGYSPQPTFRYRPTASVNRARKMADMEVANSNGLGVLTPSPQF 532
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 444 YADNRTVLNRDVEDKEFGLTTLTMNTMRDNEMAMRDMQAAAEAELEAAGFQNVRAHNG 503

QY 533 MDPLGLALHLAGTTRIGFDKATTVADNNSLVDFANLIYVAGNTIRTGFGENTPLTSMCHA 592
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 504 FAPGLGIHEMGTARMGRDPKTSVLNAHQVHECKNVYVTDGAMASASCVNPSLTMYALT 563

QY 593 IKSARSIINTLKGK 606
   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db 564 ARAADHAVRARRG 577

RESULT 7
QY9I1K8 PRELIMINARY; PRT; 591 AA.
ID Q9I1K8;
AC Q9I1K8;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
```



```
QY 187 GIGRPKLSN-----DPAEDDKENNELYSEARLIGTSTKEFDESIRHTLVLSLQDA 238
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 294 ---RQRWASEHGLSDVADPGYD---RHIDAVLERM-GVSEQCDHNGPHQ---RLVEGA 342
QY 239 YKQRQIFRPLPLACHRLKNAPEYVEHSAENLFHSIYNDQ---KOKKLEF-----286
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 343 DKLGTYTFVKA-----ALNLSPEH---YDADKAGHAGFGDGTGAKQGTINTFLKDAPEAG 393
QY 287 --LNTNHRCTRLALTGGYERKKGAAEVRNLLATRNPSQLDSYIMAKVYVILASGAIGNPQ 344
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 394 ARILVGTTRAQORVLVEDG-----RAAGYSAVTTMGDETRQIT--VRAPOVVVACGALETPA 446
QY 345 ILYNSGSGGLQVTPRNDSILNIGRYITEQPMACQIVLRQEFVDSVRDDPYG---LPWW 401
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 447 LLLRSGLGG-----PAAGRYLRLHPAGLVAGI-----YGEDQRAW 482
QY 402 KEAQAQHIKAKNPTDALPIPRDPE-----PQVTT---PFT--EEH-PWHTOI 442
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 483 GP-----PQSGILKQFADHENGHGFIIEGVQYGPALMASGLPWTGGGAHRDLMSKF 533
QY 443 HRDAFSYGAVGPEVDSRVIVD---LRWFGATDP-EANLLVFNQNDVQGYSM-----P 491
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 534 HRMATEFSIVQDRHGQVTVDEGNVAHYTALTDLDARN---FRRGVTESIRLHEAAGA 590
QY 492 OPTERYRPSTASNRARAKMMADMCEVAS---NLGGYLPSPQPMQPCGLALHLAGTTRIG 548
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 591 EEIIVALAPGVAPWRRGDDLEAFIGQVAVPLGAGGQVFS-----AHQMGSRMG 640
QY 549 FDKATTVADNNSLVYDFANLVVAGNGTIRTGFGENPILTSMCHAKSARSINTLKGGTD 608
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 641 SDPQTSVADPDGQLHDVPGVWIGDTSAPPTCSGVNPMVSMALASRTAEKLLAAWEGADG 700
QY 609 GKNTG 613
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 701 GTGSG 705

RESULT 11
Q9MOH4 PRELIMINARY; PRT; 748 AA.
AC Q9MOH4
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 82.0 KDA PROTEIN.
GN AT4G28570.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lennard N., Quail M., Harris B., Rajandream M.A., Barrell B.G.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RL EMBL; AL161573; CAB81445.1; -.
DR HSPSP; P22637; 3COX.
DR InterPro; IPR00205; NAD_binding.
KW Hypothetical protein.
SQ SEQUENCE 748 AA; 81959 MW; 13CD183F5940DAC3 CRC64;

Query Match 4.88; Score 158.5; DB 10; Length 748;
Best Local Similarity 19.48; Pred. No. 0.00058;
Matches 126; Conservative 90; Mismatches 215; Indels 217; Gaps 28;

QY 37 DVFTAGSGPTGATYAKLCVEAGLRVWVEIGAADSFTAVNAEETAVPYVPGYHKHKEIE 96
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
```

```
Db 239 DAVVVGSGGGVAAANAKAGLKVLEKG---NYFTAHDYSGLEVPMSMLEYKGG---292
QY 97 FQKDIDRFVNVYIKGALQOVSVFVRNQNVPTLDPGAWSAPPSSAISNGKNPHQREFENLS 156
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 293 -----GLLITV-----DCK-----FMLLA 306
QY 157 ABAYTRGVGGMSTHWCTGCTPRIHPMESLPGIGRPKLSNDPAEDDKENNELYSEARLIPN 216
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 307 GSAVG---GGTAVNWSAS---IRTPDHVL-----QEW---SEGSKIFK 340
QY 217 TSTKEF---DE-SIRHTLVLSLQDAVKDQRIFRPLPLACHRL-----KNAP--261
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 341 FGSQEQSAMDVEITRIGTQETRCVKHGFQ---QVLRKGCERGLGQVESVPRNSPEDH 395
QY 262 -----YVEMHSAENLFHSIYNDQKKLFTLLNHRCTRLAL-----TGGYEKK---I 306
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 396 YCGLGCGYGRAGAKNGTDQTLWLDVAVENGAVILTGKAEFRVLVDNNTSSNERKKRCYGV 455
QY 307 GAAEVRNLLATRNPSQLDSYIMAKVYVILASGAIGNPOILYNSGSGGLQVTPRNDSILPN 366
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 456 PASSVGGKGIGKK-----FIIEARVTVSSAGSLTLPPLMLSSGLKN-----PN 497
QY 367 LGR-----YITEQPMACQIVLRQEFVDSVR--DDPYGLPWWKEAVAQHIKAKN 412
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 498 IGRNLKHPVLMTWCYPPKDESEFSGKMYEGGIITSVHHMND-----TESGCKAILLEN 550
QY 413 PTDALPIPRDPEPQVTPPTTEEHPHWTQIHRDAFSGAVGPEVDSRVIVDLRWFAGADP 472
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 551 PLIG-PASYAGLSPWVS-----GPDLEKMIK-----YGR--580
QY 473 EANNLLVFQNDVQGYSM--PQPTFRYRPSTASNRARAKMMADMCEVAS---NLGGY---524
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 581 --AHLFALVRLDLSGSEVMNENEVYRTTKDRENLRAGLRVSAAGAAGEVGTYRSD 638
QY 525 -----LPTSPQEMDPGLAL-----HLACTTRIGFDKATTVADNN 559
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 639 GQMKCEAITREAMEEFLEVDVAVGGVGTGKEGYWTTYSFAHOMSGRCMGVTAEBGALDEN 698
QY 560 SLVWDFANLYVAGNGTIRTGFGENPILTSMCHAKSARSINTLKGGT 607
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 699 GESWEAEGLFVCDGSLPSAVGNPMITIQSTAVCISSKIVDSLQNK 746

RESULT 12
Q94BP3 PRELIMINARY; PRT; 748 AA.
AC Q94BP3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 82.0 KDA PROTEIN.
GN AT4G28570.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Dale J.M., Gibson H.A., Goldsmith A.D., Jiang P.X., Lee J.M.,
RA Quach H.L., Tang C., Toriumi M., Yu G., Bowser L., Carninci P.,
RA Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full Length cDNA of gene AT4g28570 (GI:7269712).";
RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
EMBL; AY039977; AAK64154.1; -.
KW Hypothetical protein.
SQ SEQUENCE 748 AA; 81987 MW; E3EB613F5D56A0D1 CRC64;
```



```
Query Match      4.8%; Score 158.5; DB 10; Length 748;
Best Local Similarity 19.4%; Pred. No. 0.00058;
Matches 126; Conservative 90; Mismatches 215; Indels 217; Gaps 28;

QY 37 DVFIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEEGTAVPVVGYHKKNEIE 96
DB 239 DAVVVGSGGGVAAANLAKGLVLEKQ---NYFTAHDYSGLEVPSMLELYEKG--- 292
QY 97 FOKDIDREVNVIKALQOQVSVPRNQNVTLPDGAWSAPPSSAISNGKNPHOREFENLS 156
DB 293 -----GLLTVV-----DQK-----FMLLA 306
QY 157 AEAVTRGVGGMSTHWTCTSPRIHPMESLPGIGRKLNSNDPAEDDKENNELYSEARLIG 216
DB 307 GSAVG---GGTAVNWSAS---IRTPDHVL-----QEW---SEGSKIIF 340
QY 217 TSTKEF---DE-SIRHTLVLSQDAYKDRQRIERPLPLACHRL-----KNAP--- 261
DB 341 FGSQEYQSAMDEVITIRIGYTERCVKHGFQ---QVLKRGGERLGLQVESVPRNSPEDH 395
QY 262 -----YVHWSAENLFHSIYNDKQKLFLLTHNRCRRLAL-----TGGYEKK-----I 306
DB 396 YCGLGCGYGRAGAKNGTDOTLWLDVAVENGAVILTGKAERFVLVDMTSSNKKKRCVGV 455
QY 307 GAAEVRNLLATRNPSQLDSYIMAKVYVILASGAIGNPQILYNSGFSGLQVTPRNDSLIPN 366
DB 456 FASSVGKGIGKK-----FIIEARVTVSSAGSLTLPMLSSGLKN-----PN 497
QY 367 LGR-----YITEQPMAFQIVLRQEFVDSVR---DDPYGLPWKKEAVAQIIAKN 412
DB 498 IGRNLKHPVLMTWGYFPEKDEFSEFGKMYEGGIITSVHMND-----TESGCKRAILEN 550
QY 413 PTDALPIPERDEPOVTTPEEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRFWGCATDP 472
DB 551 PLIG-PASTAGLSPVWS-----GFDLKERMIK-----YGR--- 580
QY 473 EANNLLVFQDQGYSM--POPTFRYRSTASNVRAKMDMCEVAS---NLGGY--- 524
DB 581 --AHLFALVRDLGSGVMNEVTVTKKDRNLRAGLRQALRSVAAGAVEVGTYRSD 638
QY 525 -----LPTSPQFMDPGLAL-----HLAGTTRIGFDKATTVADNN 559
DB 639 GOKMKCEAITKKEAMEFLEVDVAGVGVTGKEYTWTYFSAHQMGSCRMGVTAEGALDEN 698
QY 560 SLVWDFANLYVAGNGTIRTFGEFNPGLTSMCHAIRKARSIIINTLKGTT 607
DB 699 GESWEAEGFLVCDGSLPSAVGVNPMITIQSTAYCISSKIVDSLQNK 746

RESULT 13
Q988P2
ID Q988P2 PRELIMINARY; PRT; 499 AA.
AC Q988P2;
DT 01-OCT-2001 (TremBLrel. 18, Created)
DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
DE MLR6655 PROTEIN.
GN MLR6655.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idehara K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT *Complete genome structure of the nitrogen-fixing symbiotic bacterium
```

```
RT Mesorhizobium loti.*;
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003009; BAB52905.1; -.
DR InterPro; IPR00205; NAD_binding.
KW Complete proteome.
SQ SEQUENCE 499 AA; 54537 MW; 633924AED96B8F51 CRC64;

Query Match      4.7%; Score 155.5; DB 16; Length 499;
Best Local Similarity 20.7%; Pred. No. 0.00053;
Matches 129; Conservative 65; Mismatches 235; Indels 195; Gaps 27;

QY 37 DVFIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEEGTAVPVVGYHKKNEIE 96
DB 7 DIVITGSGGSSLAYSLADTGRIIVILE-----RGEHLRDTPE 45
QY 97 FOKDIDREVNVIKALQOQVSVPRNQNVTLPDGAWSAPPSSAIS-----SNCK--- 145
DB 46 ARDDIAIFQNGFYRSEE-----WLATDGESFLPGNTYYVGGNSKFFG 88
QY 146 -----NPHOREFENLSAEAVTRGVGGMSTHWTCTSPRIHP---PMESLPGIGRP 191
DB 89 AVMYRYRQEDFNPRDH-----MGGRSPQWPISYAELEPWYERAELLFGV-RG 134
QY 192 KLSNDPAEDDKENNELYSEARLIGTSTKEFDESIRHTLVLSQDAYKDRQRIERPLPL 251
DB 135 DARQDPTPEPR--NRPY---RVLPVPDPAITATVQRLLQAGIHFA-----SLPL 179
QY 252 A-----CHRLKNAPEYVHWSAENLFHSIYNDKQKLFLLTHNRCRRLALTGGYEKK 305
DB 180 AIDIDAWLRRAKTG-----WDAFPNT--GAGKIDAEVGLTKALEHPNATLITGANVORL 232
QY 306 IGAAEVRNLLATRNPSQLDSYIMAKVYVILASGAIGNPQILYNSGFSGLQVTPRNDSLIP 365
DB 233 VTDAASGRVYAAAFVFKDVELSIGADVFAAAGAVQSAALLRS-----STSVTP 282
QY 366 N-----LGR-YITEQPMAFQI-----VLRQE---FVDSVRDDPYG-LPMWKEAV 405
DB 283 NGLGNSSDQGLRNFNMHNTTAMLAIDPFRNRNTAVYQKTLGFNDYFNKDPGLSGFPLGNVOL 342
QY 406 AQHIAKNPTDA---LPIPRDEPOVTTPEEHPWHT-QIHRDAFSYGAVG---PEVD 457
DB 343 LGHITGNTILKANAPLLP-----RWLAGLVARNCYGFWFLTSEDLPNPE 384
QY 458 SRVIVDLRFWGCATDPDPAENLLVFQDQGYSMPOPTFRYRSTASNVRAKMDM--ADMC 515
DB 385 SRVTI-----RNGRIVMNVVRNMCAHETLIR-----RTRAVMREAGFP 423
QY 516 EVASNLGGYLTPSPQFMDPGLALHLAGTTRIGFDKATTVADNNSLVWDFANLYVAGNGT 575
DB 424 VVLTRTFGRKTTTS-----HQCQTARLGSDDPNTSVSPDCRSHDIANLYVTDASV 472
QY 576 IRTGFGENPTLTSCHAIKARSII 599
DB 473 LPTSAAVNPALTIATAALAKAGAAI 496

RESULT 14
O74253
ID O74253 PRELIMINARY; PRT; 769 AA.
AC O74253;
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)
DE 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE CELLOBIOSE DEHYDROGENASE (EC 1.1.99.18).
GN CDH.
OS Pycnoporus cinnabarinus.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Stereales; Schizophyllaceae; Pycnoporus.
OX NCBI_TaxID=5643;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99321795; PubMed=10393235;
```


Db 429 VVDANCKVHGMNLYVGGSSVFCTSGQANPTTT 461

Search completed: September 3, 2002, 16:00:34
Job time: 568 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 15:55:51 ; Search time 28.03 Seconds
(without alignments)
853.681 Million cell updates/sec

Title: US-09-856-327-2

Perfect score: 3284

Sequence: 1 MSLTEQMLRDPYRSMQING.....IINTLKGGTDGKNTGHEHRL 618

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	126.5	3.9	773	1	CDH_PHACH
2	116	3.5	514	1	G6PD_MYCTU
3	107.5	3.3	964	1	XINO_YEAST
4	103.5	3.2	612	1	AMTG_ASPOR
5	103.5	3.2	639	1	AMTG_ASPAK
6	102.5	3.1	1536	1	SIN3_YEAST
7	101	3.1	1010	1	SCA4_RICPA
8	100	3.0	663	1	ALOX_CANBO
9	100	3.0	1142	1	ENRM_PIG
10	98.5	3.0	382	1	MTLD_KLEPN
11	98.5	3.0	556	1	BETA_ECOLI
12	98	3.0	505	1	Y4NJ_RHISN
13	98	3.0	925	1	W70T_HUMAN
14	98	3.0	1257	1	CCAA_BACTU
15	96.5	2.9	684	1	Y492_MYCTU
16	96	2.9	291	1	US02_HSV11
17	95.5	2.9	382	1	MTLD_ECOLI
18	95.5	2.9	454	1	DLHD_RHOCA
19	95.5	2.9	519	1	AMVH_SACFI
20	95.5	2.9	708	1	HELS_SULSO
21	95.5	2.9	883	1	RPOL_BPT7
22	95	2.9	487	1	ENGA_CHLPN
23	95	2.9	500	1	YDAK_YEAST
24	94.5	2.9	855	1	GAF1_SCHPO
25	94.5	2.9	1211	1	BUN2_DROME
26	94.5	2.9	2193	1	POLG_CXAL6
27	94	2.9	606	1	PRIM_MYXXA
28	94	2.9	886	1	SMGB_MOUSE
29	93.5	2.8	406	1	YNQ5_YEAST
30	93.5	2.8	594	1	CIK1_YEAST
31	93.5	2.8	823	1	SCH9_YEAST
32	93.5	2.8	903	1	VGLB_HSV1F
33	93.5	2.8	904	1	VGLB_HSV11

34	93.5	2.8	904	1	VGLB_HSV1P
35	93.5	2.8	1859	1	RPB1_CAEEL
36	93	2.8	470	1	LEU1_AZOVI
37	93	2.8	552	1	CHOD_BREST
38	93	2.8	605	1	GOX_ASPNG
39	93	2.8	737	1	AMYL_AEDAE
40	93	2.8	890	1	GLND_ECOLI
41	93	2.8	890	1	GLND_SALTY
42	93	2.8	953	1	YNM7_YEAST
43	93	2.8	955	1	VP2_BTIV17
44	93	2.8	1012	1	UBAL_SCHPO
45	93	2.8	1117	1	CYT4_NEUCR

ALIGNMENTS

RESULT 1	CDH_PHACH	STANDARD;	PRT; 773 AA.
ID	Q01738; O00047;		
AC	30-MAY-2000 (Rel. 39, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Cellobiose dehydrogenase precursor (EC 1.1.5.1) (CDH) (cellobiose-quinone oxidoreductase).		
DE	CDH-1 AND CDH-2.		
GN	Phanerochaete chrysosporium.		
OS	Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;		
OC	Aphyllophorales; Corticiaceae; Phanerochaete.		
OX	NCBI_TaxID=5306;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=OGC101;		
RX	MEDLINE=97077226; PubMed=8919793;		
RA	Li B., Nagalla S.R., Renganathan V.;		
RT	"Cloning of a cDNA encoding cellobiose dehydrogenase, a hemoflavoenzyme from Phanerochaete chrysosporium.";		
RL	Appl. Environ. Microbiol. 62:1329-1335(1996).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=OGC101;		
RX	MEDLINE=97176414; PubMed=9023960;		
RA	Li B., Nagalla S.R., Renganathan V.;		
RT	"Cellobiose dehydrogenase from Phanerochaete chrysosporium is encoded by two allelic variants.";		
RL	Appl. Environ. Microbiol. 63:796-799(1997).		
RN	[3]		
RP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 19-208.		
RX	MEDLINE=20139694; PubMed=10673428;		
RA	Hallberg B.M., Bergfors T., Boeckbro K., Pettersson G., Henriksson G., Divine C.;		
RT	"A new scaffold for binding haem in the cytochrome domain of the extracellular flavocytochrome cellobiose dehydrogenase.";		
RL	Structure 8:79-88(2000).		
CC	- - FUNCTION: DEGRADATES BOTH LIGNIN AND CELLULOSE. OXIDIZES CELLOBIOSE TO CELLOBIONOLACTONE.		
CC	- - CATALYTIC ACTIVITY: Cellobiose + a quinone = cellobiono-1,5-lactone + a phenol.		
CC	- - COFACTOR: ONE FAD AND ONE HEME B.		
CC	- - SUBCELLULAR LOCATION: Secreted.		
CC	- - SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GMC OXIDOREDUCTASES FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; U46081; AAC49277.1; -		

DR EMBL; U65888; AAB61455.1; -;
 DR EMBL; U50409; AAB92262.1; -;
 DR PDB; 1D7B; 18-OCT-99.
 DR PDB; 1D7C; 18-OCT-99.
 DR PDB; 1D7D; 18-OCT-99.
 DR InterPro; IPR000172; GMC_oxred.
 DR InterPro; IPR001100; Pyr_redox.
 DR Pfam; PF00732; GMC_oxred; 1.
 DR PRINTS; PR00411; PNRDRTASE1.
 DR PROSITE; PS00623; GMC_OXRED_1; 1.
 DR PROSITE; PS00624; GMC_OXRED_2; 1.
 DR Cellulose degradation; Oxidoreductase; FAD; Flavoprotein; Heme;
 KW Multigene family; Signal; 3D-structure.
 FT SIGNAL 1 18
 FT CHAIN 19 773 CELLOBIOSE DEHYDROGENASE.
 FT DOMAIN 19 208 HEME DOMAIN.
 FT DOMAIN 235 773 OXIDOREDUCTASE.
 FT BINDING 83 83 HEME LIGAND.
 FT BINDING 181 181 HEME LIGAND.
 FT NP_BIND 236 265 FAD (ADP PART) (POTENTIAL).
 SQ SEQUENCE 773 AA; 82007 MW; 54F721E779AA4D7B CRC64;

Query Match 3.94; Score 126.5; DB 1; Length 773;
 Best Local Similarity 18.5%; Pred. No. 0.098;
 Matches 131; Conservative 71; Mismatches 195; Indels 313; Gaps 32;

QY 37 DVFIAGSPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEGTAV-PY----- 85
 DB 236 DYIVAGPGGIADRLSEAGKRVLLERGP-----STKGTGYVAPVAPWATSSGLTRFD 291
 QY 86 VPG-----YHKNEIEFKIDRFNVNFKGALQVQVSPVRNQVPTLDPGA--NSAPPG- 137
 DB 292 IGPLFESLFTDSNPWCKDITVFGCLVGGTSV-----GALYWPNDGD 338
 QY 138 -SSAISKGNPHQREFENLSAEATRVGVGSGTHTWCSTPRIHPMESLPGLGRPKLSND 196
 DB 339 FSSSV-----GWPSSTWTHAPYTSKLSRLPS-----TDH 368
 QY 197 PAEDDKENNEL-----YSEA-----ERLIGTSTPKFDESIRHTVLRS 234
 DB 369 PSTDQRYLESFNVVSOLLKGGYQNTINDNPNYKDHVFGYSAFDLNGKRAGPVATY 428
 QY 235 LODAYDKRQIRFRLPLACHRLKNAPEYVEHSAENLFHSIYNDKQKLFLLTNHRC 294
 DB 429 LQTA-----LARP-----NFTKTNVMS 447
 QY 295 RLALTGGYEKGIGAEVENVLLATRNPPSQLDSYIMAKV---YVLASGAINPQILYNSGF 351
 DB 448 NVVRNG---SQILG-----VQTDPTLGPNGFIPVTPKGRVILSAGAFGTSRILFQSG- 497
 QY 352 SGLQVTPRNDSLIPNLGRYITEQPMAFQIVLRQEFVDSRDDPYGLPWKEAQAHTAK 411
 DB 498 ----IGPTD-----MIQTQVS 509
 QY 412 NPTDALPIPRD-----PEQVTPPTFEHP-----WHTQIHRDAFS 448
 DB 510 NPTAAALPPQNWNLNLPVGMNAQNPNSILVFT--HPSIDAYENWADVSNRPDAAQ 567
 QY 449 YGAVGPEVDSRVIVDLRFEGATDEANNLLVFQNDVQDGYSMQPTFRY-----RPSTAS 503
 DB 568 YLANQSGV-----FAGASPKLFWRAYSG--SDGFT-----RYAQGTVRPGAAS 609
 QY 504 -----NVRAR-----KWMADNCE 516
 DB 610 VNSSLPNASQIFITVYVLTSGISGRGIGDAALRGVTLPPVLPVNPVDKTVLLQALHD 669
 QY 517 VASNLGG-----YLPFSPPQFMDPGIALHLAGTTRIGTRIGDKATTVADNN 559
 DB 670 VVSNIGSIPLGTLMTIPDVTQTLVEEVDAYDPATNSN---HWSSTTIGSSPQSAVVDN 726
 QY 560 SLVWDFANLYVAGNGTI---RTGEGENTLTSMAHKAISARSINLTGKG 606

DB 727 VKVFGTNNLFIVDAGIIPHLPFG---NPQGTLSAAEQAAKIL-ALAGG 772

RESULT 2
 G6PD_MYCTU
 ID G6PD_MYCTU STANDARD; PRT; 514 AA.
 AC O08407;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD).
 GN ZWF OR ZWF2 OR RV1447C OR MT1494 OR MTCY493.07.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J.F., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) -> D-glucono-
 CC 1,5-lactone 6-phosphate + NADPH.
 CC -1- PATHWAY: FIRST STEP IN PENTOSE PHOSPHATE PATHWAY.
 CC -1- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE
 CC FAMILY.
 CC -1- CAUTION: M.TUBERCULOSIS HAS TWO GENES FOR ZWF. THIS ONE LOOKS LIKE
 CC A CLASSICAL ZWF.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z95844; CAB09259.1; -;
 DR EMBL; AE007019; AAK45757.1; -;
 DR HSP; P11411; IDPG.
 DR TIGR; MT1494; -;
 DR TubercuList; RV1447C; -;
 DR InterPro; IPR001282; G6PD.
 DR Pfam; PF00479; G6PD; 1.
 DR Pfam; PF02781; G6PD_C; 1.
 DR PRINTS; PR00079; G6PDHDRGNASE.
 DR PRODOM; PD001129; G6PD; 1.
 DR PROSITE; PS00069; G6P_DEHYDROGENASE;
 DR Oxidoreductase; NADP; Glucose metabolism; Complete proteome.
 KW ACT SITE 210 210 BY SIMILARITY
 FT ACT SITE 210 210
 SQ SEQUENCE 514 AA; 57343 MW; 22A9CCDC9AB062F CRC64;

Query Match 3.5%; Score 116; DB 1; Length 514;
 Best Local Similarity 22.8%; Pred. No. 0.34;
 Matches 98; Conservative 48; Mismatches 163; Indels 120; Gaps 23;

```

QY 104 FVNYIKALQOQ-VSVYVRNQNVPTLPDGAASAPPGSSNAISGNKPNHQREFENLSAEAVTR 162
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 77 FGQVYNAVGEHCRTFPRQQWDRLAG-PRFVPGTFDDDDAFQAQLAETLEKLDAAE---R 132
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 163 GVGGMSTHWTCSTPRIHPPM-ESL--PCIGRPKLSNDPAEDDKENNELYSAEERLIGTST 219
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 133 GTGGNHAFYLAIPKSPFVCEQLHKSGSLARPQ-----GDRSRVIVIE----- 175
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 220 KEFDESTRHVLVLSRLQADYKDRQIRFRPLPLACHRL--KNAPEYVWHSNENLFHSIYN 277
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 176 KPFQHDLASARELNKANVAVPEEVAER---IDHLYGKETVQVILALRFANQLDFPIWN 231
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 278 ----DDQKKLFTLTNHRCTRRLALTGGYEKKIGAAE--VRN-----LLATRNPSQL 324
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 232 AHYVDHVQ-----ITWAEIDIGLAGRAGYDGIAGARDVIONHLMQLLALTAMEEVSFH 285
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 325 DSYIMA-KVYVLASGAIGNP-----QILYNSGFSG----- 353
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 286 PAALQAEKIKVLSATRLAEPLDQTTSRQYAAAGWQGEKVYGLLDEEGFAEDSTTETFAA 345
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 354 --LOVTRNDSLIP-----NLGRYITEQPMATFCQIVLRQEFV--DSVRDDPYGLPWKK 402
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 346 ITLEVDTRRWAGVDFYILRTGKRIGRRVTEALVF-----RRAPHLPFDAATMDELG----- 396
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 403 EAVAQHIAKNPDTALPIPFPRPEQVPTPTFTTEHPHWTQIHRDA---FSYGAV---GPE 455
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 397 -----TNAMVIRVQ-PDEGVILRGSKYPGTGAMEVDYNNMDFSYGSAFAEDSPE 444
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 456 VDSRVIVDL 464
      | : | : |
DB 445 AYERLILDV 453
      | : | : |

RESULT 3
YIN0_YEAST
ID YIN0_YEAST STANDARD; PRT; 964 AA.
AC P40467;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative 108.8 kDa transcriptional regulatory protein in FKHL-STH1
DE intergenic region.
GN YIL130W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId:4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moulie S., Moulie T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: CONTAINS 1 ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR
CC CLUSTER DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; Z38059; CAA86148.1; -.
DR PIR; S48404; S48404.
DR HSP; P08657; ICLD.
DR SGD; S0001392; YIL130W.
DR InterPro; IPRO01138; zn2_cy6_fungal.
DR Pfam; PF00172; Zn_Clus; 1.
DR PRINTS; PR00054; FUNGALZNCYS.
DR SMART; SM00066; GAL4; 1.
DR PROSITE; PS00463; ZN2_CY6_FUNGAL_1; 1.
DR PROSITE; PS50048; ZN2_CY6_FUNGAL_2; 1.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Nuclear protein; Zinc; Metal-binding.
FT DNAS_BIND 21 47      ZN(2)-CYS(6), FUNGAL-TYPE.
FT DOMAIN 811 896     HIS-RICH.
SQ SEQUENCE 964 AA; 108780 MW; AD5ABE59E4B022CC CRC64;

Query Match          3.3%; Score 107.5; DB 1; Length 964;
Best Local Similarity 19.5%; Pred. No. 3.8;
Matches 113; Conservative 86; Mismatches 187; Indels 193; Gaps 30;

QY 33 NGDVDFITAGSGPIGANVAK--LCVEAGLRVVMVEIGAADSFYAVNAEEGTAVPYV--- 86
DB 322 DEGYKYFTAARKLLIDITNARDLNSIQAILMLII-----FLOCSARLSCTCYIIGVAM 373

QY 87 -----PGVHKK-----NEIEFOK-----DIDRFVNVIKGALQQOVSVPRNNQVP 125
DB 374 RSALARCFHRKLSPNSGFSPTIEEMRKLFYYIKLDVINAMGLPRSISPDDEFDTLP 433

QY 126 TLD-----PGAWSAPPSSAISGNKPQHREF-----ENLSAEAVTRGVGGMST 169
DB 434 -LDLSDENITEVAYLPENOHSLVSTGIS---NEHTKLFLLINELISELYPIKKTSLNIS 489

QY 170 HWTCST-PRTHPMESLPGICRPKLSN-DPAEDDKEWNELYSAERLIGTSTREFDESI 226
DB 490 HETVTSLELKLRNLWDSLPELIPNAENIDPE-----YERANRLLHL-----SF 533

QY 227 RHTVLRLSLQDAYKDRQRIFRPL-----LYPFFHYLSRNMAENVDPCLYRRARNSTAVARTVIKLAKE 581
DB 534 LHVQLII-----WHSAENLFHSIY-----NDKQKKLFTLLTNHRCTRALTGG 301

QY 262 YVE-----HWSAENLFHSIY-----NDKQKKLFTLLTNHRCTRALTGG 301
DB 582 MVSNLLLTGSYWAYCYTIIFYSVAGLLFYIHEAQDPDKDSAREYVDILKDAETGRSVLIQL 641

QY 302 YEKKIGAEVENLNLATRNPSQSOLDSYIMAKYVVIVLASGAIGNPQILYN--SCFSGLQVTPR 359
DB 642 KDSMSASRTYNLL-----NQIFEKLSNTKIQTJA-----LHSSPSNESAFLVTTNN 687

QY 360 NDSLIPNLGRYITQPMAFCQIVLRQEF-----VDSVRDDPYGLPWKKEVAOHIKAKNPT 414
DB 688 SSALKPHLGDSL-QPPVFESSQDTKNFSIAKSESIND-----YAMANYLNWTPI 737

QY 415 DALPIPFDRPEPQV---TTPTEEHHPWHTQIHRDAFSYGAVGEVDSRVIVDLRWFGATD 471
DB 738 SENPLNEAQQOVQSQGTNNMSNE-----RDPNNFLSIDIRLDNN----- 777

QY 472 PEANNLLVFQNDV---QDGYSNMPOQT---FYRPFSTASN 504
DB 778 -GQSNILDATDDVFIRNDG-DIPTNSAFDFSSSKSNASN 814

RESULT 4
ID AMYG_ASPOR STANDARD; PERT; 612 AA.
AC P36914;
DT 01-JUN-1994 (Rel. 29, Created)
DE 01-JUN-1994 (Rel. 29, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Glucoamylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
DE (N.1.4-alpha-D-glucan glucohydrolase).
GN GLAA.
```


between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; D00427; BRA00331.1; --
PIR; J0479; J0479.
HSP; P04064; IGAI.
GlycoSuiteDB; P23176; --
InterPro; IPR002044; CBD.4.
InterPro; IPR000165; Glyco_hydro_15.
Pfam; PF00686; CBD_4; 1.
Pfam; PF00723; Glyco_hydro_15; 1.
PRINTS; PR00736; GLYDRLASE1.
ProDom; PD001568; CBD_4; 1.
ProSITE; PS00820; GLUCOAMYLASE; 1.
Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
Signal.
KW SIGNAL 1 18
FT PROPEP 19 24
FT CHAIN 25 639
FT DOMAIN 494 538
FT BINDING 143 143
FT ACT_SITE 199 199
FT ACT_SITE 202 202
FT ACT_SITE 203 203
FT DISULFID 233 236
FT DISULFID 245 472
FT DISULFID 285 293
FT DISULFID 194 194
FT CARBOHYD 418 418
FT CARBOHYD 464 464
FT CARBOHYD 466 466
FT CARBOHYD 467 467
FT CARBOHYD 475 475
FT CARBOHYD 476 476
FT CARBOHYD 482 482
FT CARBOHYD 483 483
FT CARBOHYD 485 485
FT CARBOHYD 487 487
FT CARBOHYD 491 491
FT CARBOHYD 495 495
FT CARBOHYD 498 498
FT CARBOHYD 499 499
FT CARBOHYD 500 500
FT CARBOHYD 501 501
FT CARBOHYD 503 503
FT CARBOHYD 505 505
FT CARBOHYD 507 507
FT CARBOHYD 511 511
FT CARBOHYD 512 512
FT CARBOHYD 513 513
FT CARBOHYD 514 514
FT CARBOHYD 516 516
FT CARBOHYD 517 517
FT CARBOHYD 519 519
FT CARBOHYD 521 521
FT CARBOHYD 523 523
FT CARBOHYD 524 524
FT CARBOHYD 525 525
FT CARBOHYD 526 526
FT CARBOHYD 527 527
FT CARBOHYD 528 528
FT CARBOHYD 529 529
FT CARBOHYD 530 530
FT CARBOHYD 531 531
FT CARBOHYD 533 533
FT CARBOHYD 534 534
KW SEQUENCE 639 AA; 68271 MW; E112B31A4DD8DD6B CRC64;

Query Match 3.2%; Score 103.5; DB 1; Length 639;
Best Local Similarity 18.1%; Pred. No. 4.2;
Matches 119; Conservative 68; Mismatches 233; Indels 239; Gaps 28;
QY 22 IPKNAIHEITGNDGVDFIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFAVNAEET 81
DB 37 VARTAILNNTIGADG--AWVSGA-----DSGIVASPTDNDPDYFTYTRSG 82
QY 82 AVPYVPGYHKKKEIEFQKIDRFVNVKIGALQQVSVPRVNONVPTLDPGAWSA 141
DB 83 VIKTLVDLFRNGDTDLLSTIEHYI-----SSGILGEPEKFNVDYETAYT 106
QY 142 SNKNPHQREFEENLSAEAVTRGVCGMSTHTWCSTPRTHPPMESLPGICRPKLSNDPAEDD 201
DB 107 -----SSQAIIGVSNPSGDL-----SSGILGEPEKFNVDYETAYT 140
QY 202 KENNELYSEARLIGTSTKEFDESIRHTLVLSLQDAYK--RQIFRPLPLACHRLKNAP 260
DB 141 GSGRPQPDGALRATAMIGFQVL-----LDNGYTSAAETIVWPL-----VRNDL 186
QY 261 EYVE--WHSANLPHSYNDKQKFLTLNHRCTRRLALTGG--YEKKIGAAEVRNLLA 316
DB 187 SYVAQYWNQTG---YDLWEENVSGSFFETIAVOHR---ALVEGSAFATAVGSS----- 232
QY 317 TRNPSSQLDSYIMAKVYVLASCAIGNQOIL--YNSGFSGLQVTPRNDSLIPN-----LG 368
DB 233 ----CSWCDQA-----FOILCYLQSFWTGSIYILANFDSRRSGKDTNTLLG 274
QY 369 RYIEQPMAFQ-----IVLRQEFVDSYR-----DDPY 396
DB 275 SIHTFDEACGDDSTFQPCSPRALNKHKEVVDSPRSYITLNDGLSDSEAVAGVPEDSY 334
QY 397 --GLPWPKEAQAQIAKNPTDPIPRDEPOVTPFTTEHPWHTQ-----IH 443
DB 335 YNGNPWFQSTLA--AAEQLYDAL-----YQDKQGSLEITDVSLEDF 374
QY 444 RDAFSYCAVG-----PEVDSRVIVDLRWFG-----ATDPEANLLVFQNDVQDG--YS 489
DB 375 KALYSGAATGTYSSSSSYSSIVSAVKTFADGFGYSIVETHAASNGSLSEQDKSDGDEL 434
QY 490 MPOPTFRYRPSSTASNVARKMM-----ADCEVASNLGGY---LPTSPQPM 534
DB 435 ARDLTWSYAALLTANNRNSVVPFSGWETSSASWPGTCATASGTSYSSVTTSWPSIVA 494
QY 535 PGLALHLAGTTRIG----FDKATTVADNNSLVWDFAN-----LYVAGNGTIRTFGFGN 583
DB 495 TGGTTTATTGGGVTSTKSTTTTASKTSTTSSSTCTTPTTAVAVTFDLTATTYGEN 553
RESULT 6
SIN3_YEAST
ID SIN3_YEAST STANDARD; PRT; 1536 AA.
AC P22579; Q08049;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Paired amphipathic helix protein.
GN SIN3 OR SDI1 OR UME4 OR RPD1 OR GAM2 OR SDS16 OR YOL004W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=GRF8;
RX MEDLINE=91042523; PubMed=2233725;
RA Wang H., Clark I., Nicholson P.R., Herskowitz I., Stillman D.J.;
RT "The Saccharomyces cerevisiae SIN3 gene, a negative regulator of HO,
RL contains four paired amphipathic helix motifs.";
RN Mol. Cell. Biol. 10:5927-5936(1990).
RP [2]
RA SEQUENCE FROM N.A.
RA Hughes B., Pohl T.M.;

Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC - FUNCTION: REQUIRED FOR BOTH FULL TRANSCRIPTION REPRESSION AND
 CC ACTIVATION OF MANY GENES INCLUDING CELL TYPE-SPECIFIC GENES (STE6,
 CC TY2 AND HO), CELL DIFFERENTIATION-SPECIFIC GENES (SP013), GENES
 CC THAT RESPOND TO EXTERNAL SIGNALS (PHO5) AND TRK2. IT IS PROBABLY
 CC INVOLVED IN THE SAME TRANSCRIPTIONAL REGULATORY FUNCTION OR
 CC PATHWAY AS THE TRANSCRIPTIONAL REGULATORY PROTEIN RPD3.
 CC - SUBUNIT: IT PROBABLY FORMS A COMPLEX WITH THE TRANSCRIPTIONAL
 CC REGULATORY PROTEIN RPD3.
 CC - SUBCELLULAR LOCATION: NUCLEAR, POSSIBLY LINKED TO CENTROMERE.
 CC - DOMAIN: CONTAINS 4 PAIRED AMPHIPATHIC HELIX MOTIFS, SEPARATED BY A
 CC 10 TO 30 AA SEGMENT THAT FORMS POSSIBLY A LOOP, RESULTING IN A
 CC STRUCTURE SIMILAR TO THAT OF HLH AND TPR MOTIFS.
 CC - SIMILARITY: TO S.POMBE SPAC12C2.10C.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; M36822; AAA34839.1; -
 DR EMBL; 274746; CAA99003.1; -
 DR PIR; S12068; RGBYS3.
 DR SGD; S0005364; SIN3.
 DR InterPro; IPR003822; PAH.
 DR Pfam; PF02671; PAH; 3.
 KW Repeat; Transcription regulation; Repressor; Cell division;
 KW Activator; Nuclear protein.
 FT DOMAIN 239 1200 4 X 2 PAIRED (A,B) AMPHIPATHIC HELICES.
 FT REPEAT 239 252
 FT REPEAT 274 285
 FT REPEAT 426 439
 FT REPEAT 461 472
 FT REPEAT 679 692
 FT REPEAT 714 725
 FT REPEAT 1152 1165
 FT REPEAT 1189 1200
 FT DOMAIN 480 519 GLN-RICH
 FT CONFLICT 510 510 Q -> QAO (IN REF. 2).
 SQ SEQUENCE 1536 AA; 174838 MW; 0834726312B13878 CRC64;

Query Match 3.1%; Score 102.5; DB 1; Length 1536;
 Best Local Similarity 19.3%; Pred. No. 19;
 Matches 115; Conservative 83; Mismatches 192; Indels 207; Gaps 31;
 QY 72 FYAVNAEGTAVYVPGYHKKNEIEFQKIDIRFVNV---IKGALQGVSVPRNQNVPTLD 128
 DB 985 FYDILCLADFTFHTTAYSNPKERLKLKLVYFISLFFSIFKTEESLYSHKQNV--- 1040
 QY 129 PGAWSAPGS---SALSNCKNPHOREFENLSA-----EAVTRGVGGMSTHTCTSPRTH 179
 DB 1041 ----SSSGSDGSSTASRKYQEQEMSLDILHRSRYQKLKR-----SNDEGKVPQLS 1091
 QY 180 PMWESLPG-IGRPKLSNDPAEDD-----KEM 204
 DB 1092 EPPEEPNTIEELIDEAKNPWLGNLVEANSOGIIONRSIFNLFANTNIYFFRW 1151
 QY 205 NELYSEAEKLTGTSTKEFDESIRHTLVLRSL-----QDAYKD 241
 DB 1152 TTIY---ERLL--ETQMNERTVKEINTRSTVTFAKDLDLLSSQLSEMCLDFVGDYAKQ 1206
 QY 242 RORIFRPLPLACHRLKNAPEYVWEHSAENLFHSIYNDKQKLFYL-----LTHRCRTR 295
 DB 1207 VLRLSR-----RLNGDLEHQWF--EESLRQAVN-NKAFKLYTIDKVTQSLVKAHTP- 1255
 QY 296 LALTGGYEKKGAEEVRNLLATRNPSQLDSVIMAKVYVVLASGACIGNPOILYNSGFSGLQ 355
 DB 1256 -LMTDAKTAIMEALFVKD-----RN-----ASTSAKOQIIYR-----LQ 1289

QY 356 VTFRNDSLIPNLGR-----YITEQPMACQIVLROEFVDSVRDDPYGLPWKEAVQA 407
 DB 1290 VR-SHMSNTENNMFIEFKRTRLHVSIOYIALDDLTLKPKADEK-----WKYYVTS 1340
 QY 408 HTAKNPTDALPFPDRPEQVTPTEEPHWHTQHRODAFSYCAVGPVDSRVIVDLRWF 467
 DB 1341 YALPHPTGEG-----PHEKLKIPFE-----RLIEFGQDIDGTEYDEF----- 1379
 QY 468 GATDPEANNL-----LVFQNDVQDG-YSM--PQTFERYRPSSTASNVARKMADMC 515
 DB 1380 ---SPEGISVSLTKIKIPIYQLHIENGSDYVFRKATNKY-PTIANDTQKGHVSKK 1435
 QY 516 EVASNLGGYLPSPQFMD--PGLALHLAGTTRIGFDK-----ATTVDNNSL 561
 DB 1436 ELIS-----KFLDCAVGLRNLDLDEAQLSKMOKKWNKLSIAKTSAGNQGI 1481
 RESULT 7
 ID SCA4_RICPA STANDARD; PRT: 1010 AA.
 AC Q9AJ75;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
 DE (Protein PS 120) (Fragment).
 GN SCA4 OR D.
 OS Rickettsia parkeri.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_TaxID=35792;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sekeyova Z., Roux V., Raoult D.;
 RT "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
 RT 'gene D' coding for an intracytoplasmic protein.";
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 CC - SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; AF155059; AAK30690.1; -
 KW Antigen.
 FT NON_TER 1 1
 FT NON_TER 1010 1010
 SQ SEQUENCE 1010 AA; 110694 MW; 78586D8C92FF9C5B CRC64;

Query Match 3.1%; Score 101; DB 1; Length 1010;
 Best Local Similarity 18.2%; Pred. No. 13;
 Matches 106; Conservative 88; Mismatches 243; Indels 144; Gaps 25;

QY 136 PGSSAISNCKNPHOREFENLSAEAVTRGVGGMSTHTCTSPR----IHPPMESLPGIGRP 191
 DB 55 PMSVLSGNISPSQ-----TSDPITKAV-----RETIQPKQKDLIEQLDLAAL---- 100
 QY 192 KLSNDPAEDDKENNELYSEARLIGT-----SKTEF-DESIHRTLVLRSLQ-----AYKD 241
 DB 101 -TRDLAEQKREIEEKEKDKTLSTFFCNPANREIDKALENPELKKLESIEIAGYKN 159
 QY 242 RORIFRPLPLACHRLKNAPEYVWEH---SAENLFHSIYNDKQKLFYL----- 287
 DB 160 VHNTES---AASGYGGGFKPVQWENHVSASDLRATVVKNDAGDELCTLNNTVTKPPT 215
 QY 288 LTNHRCRTRALTGGYE-----KKIGAEVRNLLATR---NPSSQLDSVIMAKVYVL 335
 DB 216 LAKQDGTQVQISSYREIDFPPIKLDKADGSMHL-SMVALKADGTKPSKOPVYFAH---Y 271

RL Adv. Dent. Res. 10:111-118(1996).
CC -!- FUNCTION: PEPTIDES DERIVED FROM THE PARENT ENAMELIN ARE COMPONENTS
CC OF ENAMEL, A UNIQUE AND HIGHLY MINERALIZED ECTODERMAL TISSUE
CC COVERING VERTEBRATE TEETH.
CC -!- TISSUE SPECIFICITY: EXPRESSED BY SECRETORY-PHASE AMELOBLASTS.
CC INTACT ENAMELIN AND LARGE-MOLECULAR-WEIGHT ENAMELINS ARE LIMITED
CC TO THE MOST SUPERFICIAL LAYER OF THE DEVELOPING ENAMEL MATRIX,
CC WHILE LOW-MOLECULAR-WEIGHT ENAMELINS ARE OBSERVED IN DEEPER
CC ENAMELIN. PREFERENTIAL LOCALIZATION AMONG THE CRYSTALLITES IN ROD
CC AND INTERROD ENAMEL.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED FROM LATE DIFFERENTIATION TO THE
CC TRANSITION STAGE.
CC -!- PTM: PARENT ENAMELIN IS PROTEOLYTICALLY CLEAVED INTO SEVERAL
CC SMALLER POLYPEPTIDES. CLEAVAGE OF N-TERMINAL REGION OF ENAMELIN
CC OCCURS SOON AFTER SECRETION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U52196; AAD10837.1; -
DR GlycoSuitDB; O97939; -
KW Signal; Enamel; Glycoprotein; Hydroxylation; Phosphorylation.
FT SIGNAL 1 38
FT CHAIN 39 1142 ENAMELIN
FT CHAIN 39 2 56 KDA ENAMELIN.
FT CHAIN 39 665 89 KDA ENAMELIN.
FT CHAIN 39 142 KDA ENAMELIN.
FT CHAIN 39 155 KDA ENAMELIN.
FT CHAIN 174 276 32 KDA ENAMELIN.
FT CHAIN 515 665 25 KDA ENAMELIN.
FT CHAIN 670 34 KDA ENAMELIN.
FT CHAIN ? 45 KDA ENAMELIN.
FT MOD_RES 53 53 PHOSPHORYLATION (PROBABLE).
FT MOD_RES 191 191 PHOSPHORYLATION.
FT MOD_RES 216 216 PHOSPHORYLATION.
FT MOD_RES 547 547 PHOSPHORYLATION.
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .).
FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .).
FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 929 929 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1040 1040 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 680 680 H -> D (IN REF. 2).
FT CONFLICT 838 840 RDH -> TTI (IN REF. 2).
SQ SEQUENCE 1142 AA; 128352 MW; 938306BC87CC5FC6 CRC64;

Query Match 3.0%; Score 100; DB 1; Length 1142;
Best Local Similarity 19.2%; Pred. No. 18;
Matches 118; Conservative 65; Mismatches 210; Indels 220; Gaps 32;

QY 76 NAEETGAVPYVP--GVH-----KKNEFEOKDIDR-----FVNVIKG 110
DB 192 NEEGNN--PYGFFGHHGGRRPPYSEMFQDEKPEKDPKTPATPSVN----245
QY 111 ALQGVSVVRNVPTL-DP-GAWSPAGSSAISGNKPNQR-----EFENLSAEAVTR 162
DB 246 ----TTVPETNTQPNAPNPRGNDTSPGTGS--GQPNPRSNPTGQNGPVAVNSGGVPR 299
QY 163 -----GVGGMSTHTWCSTPRI-----HPPMELSLPGIGR-----PKLSNDPAEDDKW 204
DB 300 SOSPPGPRQTIITHENYPNIRGFFPARRQWRPFG---PAMGHRNRGPFYRNQOIQRGPRW 356
QY 205 NELYSEARLICTSKEFDESIRHTLVLSQDAYKDRIERPLPLACHRLKNAPEYVE 264
DB 357 NSFTELEK-----QAVRPGPTTTRVY-----GSTARSNPNTYA- 390

QY 265 WHSAENLFHSIYNDKQKKLFTLLTNHRCRIALATG-----GYEKKTGAEE 310
DB 391 -GNSANLRKPEGNK-----NPMVTNVAPPKPGKGTVDQENENIQPREKQVSQKE 440
QY 311 VRNLLATRNPS-----SQDSYIMAKYVYVLASGAIGNPQILYNSGFSGLQVTPRNDSLIPN 366
DB 441 -RTVVTPTDPSGFWNRNSQDYGKSNYKL-----PQ-----PEDNMLVPN 479
QY 367 LG-----RYITEQPMAFQIIVLRQEFVDSVRDDPYGLPWKEA---VAQ 407
DB 480 FNSIDORENSYPRGESKRAPNSDGTQTQIIPK-----GIVLEPRRIPESETNOPELK 534
QY 408 HIAKNP--TDALIPFRDPEP-----QVTPTEEHPMHTQIHRDAFSYGAVGPEV 456
DB 535 HSAIQPVYTEGIPSPAKHFPAGRNWTNQOEISPPPEKED-PGRQEBHLPHLSHG-----587
QY 457 DSRVIVDLRFWFGATDPEANLLVFQNDVODGYSMP-----QPTFRYRPSSTASNVARKMMA 512
DB 588 -SRVHYYPDPYDPRENSPYLSRNTWYERDDSPNTMGQPNPHYPMTDPDKET----642
QY 513 DMCEVASNLGGVLPSPQPMFMDPLGLALHLAGTTRIGFDKATTVADNNSLVWDFANLYVAG 572
DB 643 -----IPNEEDPIDTGDGHEFPQOSR-----WDMEELSFKE 674
QY 573 NCTIRTFGENPT 585
DB 675 DPTVRHYEGEQYT 687

RESULT 10
MTLD_KLEPN STANDARD; PRT; 382 AA.
AC Q9XB6;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mannitol-1-phosphate 5-dehydrogenase (EC 1.1.1.17).
GN MTLD.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KAY 2026;
RA Otte S., Lengler J.W.;
RT "The mtl genes and the mannitol-1-phosphate-dehydrogenase from
RT Klebsiella pneumoniae."
RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC -!- CATALYTIC ACTIVITY: D-mannitol 1-phosphate + NAD(+) = D-fructose
CC 6-phosphate + NADH.
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF166095; AAD45386.1; -
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDDHGRNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 3 14
SQ SEQUENCE 382 AA; 41130 MW; D9927B6C068C5344 CRC64;

Query Match 3.0%; Score 98.5; DB 1; Length 382;
Best Local Similarity 22.0%; Pred. No. 4.6;

Db 108 PPGQALPSAPGVVLG--PEDLPVEVLQFHTSDGILVSAAGTIVKVDAAKQ---PLT 162
Qy 184 SLPGIGRKLNDPAEDDKKENNELYSSEARLIGTSTKEFDESIRHTLVLRSLQDAYKDRQ 243
Db 163 EL-----AAHGLVQSAVMSRDGALVGT-----ACKDKQ 191
Qy 244 -RIRPLPLACHRLK-NAPEYVEVH-----SAENLFHSIYNDKOKKL----- 284
Db 192 LRIEDP-----RTKPRASQSTOAHENSRRSLRMAWGTWEHLVSTGFNQMREREYKLMWT 245
Qy 285 -----FTLLNHRT-----RLALTGYEKKIGAAEVRNLLAARNFSSQ--L 324
Db 246 RFSSALASLTLDISLGLVPLLPDPSGLVLVLAGGERQLCYEVVPOOPALSPVTCVL 305
Qy 325 DSYIMAKVYVILASGAIGNPQILYNSGSLGQVTPRNDLSLPLNGLRYITEQPMACQIVLR 384
Db 306 ES-----VLRGAAVPRQALVAMGCEVLRVQLSDTAIVPIGYHVPRKAVEFHEDL-- 356
Qy 385 QEFVDSV-----RDDPYGLPWW--KEAVAQHIKAP-----TDALPIPRDPPEQVTT 430
Db 357 --FPDTAGCVPATPHG--WAGDNQQVQKVSINPACRPHPSFTSCL-VPPAEPLPDTAQ 411
Qy 431 PFTEHPWHTQIHRDAFSYGAVGEVDSRVIVDLRWFGATPEANLLVFNQDVGYSM 490
Db 412 PAVMETP-----VGDADASEGFS 430
Qy 491 PQPTFRYRPSPASNV 505
Db 431 P-PSLTSPTSPSSL 444

RESULT 14

ID CCAA_BACTO STANDARD; PRT: 1257 AA.
AC Q45754;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pesticidal crystal protein cry2Aa (Insecticidal delta-endotoxin
DE CryIIIA(a)) (Crystalline entomocidal protoxin) (142 kDa crystal
DE protein).
GN Cry12AA OR CRYIIIA(A) OR CRYVB.
OS Bacillus thuringiensis
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1428;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 2-11.
RC STRAIN=NRRL B-18244 / PS33F2;
RA Narva K.E., Payne J.M., Schwab G.E., Hickie L.A., Galasan T.,
RA Sick A.J.;
RT "Novel Bacillus thuringiensis microbes active against nematodes, and
RT genes encoding novel nematode-active toxins cloned from Bacillus
RT thuringi.";
RL Patent number EP0462721, 27-DEC-1991.
CC -1- FUNCTION: ENDOTOXIN WITH NEMATOCIDAL ACTIVITY.
CC -1- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING
CC SPORULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART
CC OF THE SPORE COAT.
CC -1- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
CC N-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L07027; AAA22355.1; -
DR InterPro: IPR001178; Endotoxin.

DR Pfam: PF00555; endotoxin; 2.
RW Toxin; Sporulation.
SQ SEQUENCE 1257 AA; 142265 MW; 3D9888FEC6C0E3981 CRC64;
Query Match 3.0%; Score 98; DB 1; Length 1257;
Best Local Similarity 20.3%; Pred. No. 30;
Matches 132; Conservative 88; Mismatches 243; Indels 188; Gaps 36;
Qy 15 SMQINGQIPKNAIHETHTYGNQGVDFIAGSGPIGATYAKLVCEAGLRVVMVEIGAADSFA 74
Db 300 SSPIYQVPVKMNMQNTSSISIVPSDLFHYQGLVKLEFSTRTDNDGLAKFT--GIINTFYK 357
Qy 75 VNAEGTAVPVVPGYHKKNEIEFQKIDRFNVVIGALQOVSVVPRNVNPT--LDPGAMS 133
Db 358 -----SPNTHETHYHDFSYNTQSSNISRSGSNPITDLNPNPIISTCIRNFEYK 406
Qy 134 APPGSSAISNCKNPHOREFENLSAEAVTRGVCGNTHWTCSTPRHPHWPESLPGICRKL 193
Db 407 AIAGSSVLVNFKDGTO---GYAFAQAPTGA-----WD-----HSFIES----- 442
Qy 194 SNDAEODDKENNELYSSEARLIGTSTKEFDESIRHTLV--LRSLQDAYKDRIRPLPLA 252
Db 443 --DCAPEGHKLNITYTSP---GDTLRDFINV--YTLISTPTINELSTEKIKGP---PAE 491
Qy 253 CHRLKN-----APEYVEHSAENLFHSIYNDKOKKFLTLTNHRCRTRALTTGGYEK 304
Db 492 KGYIKNOGIMKYGKPEYINGAQPVL-----ENQOTLIFE-----HASKTAQVTI 538
Qy 305 KIGAAEVRNL-----LATRNPSSQLDSYIMAKVYVILASGAIGNPQILYNSGFSGL 354
Db 539 RIRYASTQGTGKGYFRLDNQELQTLNPTSHNGYV-----TGNIGENYDLYTIG--SY 588
Qy 355 QVTPRNDLSLPLNGLRYITEQPMACQI-----VL-ROEFV--DSVRDOPYGLPWKE 403
Db 589 TITEGNHTL-----QIOHNDKNGVLDRIEFPVKDSQDSLP----- 624
Qy 404 AVAQHIANKPTDALPIPRDPPEQV---TTPFTEHP--W-----HTQIHRDAFSYGAV 452
Db 625 -----QDSPPEVHSTIIFDKSSPTIWSNKHSHYSHILEG--SYTSQ 665
Qy 453 GPEVDSRVIVDLRWFGATDPEANLLVFON--DVQDGYSMPOPTFYRSTASNVPRKMM 511
Db 666 G-SYPHNLLINL--FHPTDPNRNHTIHNNGMDMNVYG-----KDSVADGLNFNKIT 714
Qy 512 ADMCEVASNLG-----GYLPTSPQFMDPGIALHLAG--TTRIGFDKAT-----TVA 556
Db 715 ATIPSDAWYSGTITSMHLFNDNNFKTITPKF---ELSELENIITTOVNAFASSAQDTLA 771
Qy 557 DNNSLVWDFANLYVAGNGTIRTFG--ENPTLTSMCHAIKSARSIIINTLKG 606
Db 772 SNVSDYW-IEQVMKVDALSDEVFGKEKKALKRLVQAKRLSKIRNLLIG 821
RESULT 15
Y492_MYCTU
ID Y492_MYCTU STANDARD; PRT: 684 AA.
AC Q11157;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 57.3 kDa protein GMC-type oxidoreductase RV0492c.
GN RV0492C OR MT0511/MT0512 OR MTCY20G9.18C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Gordon S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,

RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.*;
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN= CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.*;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RC -!- COFACTOR: FAD (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE GMC OXIDOREDUCTASES FAMILY.
CC -!- CAUTION: REF.2 SEQUENCES DIFFER FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 196.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 277162; CAB00953.1; .
DR EMBL; AE006952; AAK44734.1; ALT_FRAME.
DR EMBL; AE006952; AAK44735.1; ALT_FRAME.
DR TIGR; MT0511; .
DR TIGR; MT0512; .
DR Tuberculist; RV0492c; .
DR InterPro; IPR001450; 4Fe4S_ferredoxin.
DR InterPro; IPR000172; GMC_oxred.
DR InterPro; IPR00205; NAD_binding.
DR Pfam; PF00732; GMC_oxred; 1.
DR PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
DR PROSITE; PS00624; GMC_OXRED_2; 1.
KW Hypothetical protein; Oxidoreductase; Flavoprotein; FAD;
KW Complete proteome.
SQ SEQUENCE 684 AA; 72253 MW; 0242535D29E9FB7B2 CRC64;

Query Match 2.9%; Score 96.5; DB 1; Length 684;
Best Local Similarity 19.2%; Pred. No. 16;
Matches 123; Conservative 61; Mismatches 224; Indels 233; Gaps 27;
QY 37 DVFIAAGSPIGATYAKLVEAGLRVVMVEIGAADSEFYAVNAEEGTAVPVPGYHKHNEIE 96
DB 203 DAVVVGSGAGGAMVARTLARAGLDVVLVEGRR---WTV-----EEFR 242
QY 97 FQKIDRFVNVTKALQOVSPVRNQNVPTLDPGAWSPGSSAISNGKNPKHQREFENLS 156
DB 243 STHPVDRVAGLYRGAGATVA-----LGRPAVVLPMG----- 273
QY 157 ABAVTRGVGNGS--THWTCSTPRIHPHMESLPGIKPLSNDPAEDKWNELYSAEARL 214
DB 274 -----RAVGGTTVVNSGTCFRPSL-----AVQRRWRDEGLGLADPDQLGRRL--DDAEGT 322
QY 215 IGTSTKEFDESIRHVLVLSLDQAYKDRQIRPLPLACHRLKNAP----- 260
DB 323 LRVAVPVPLEIMGRNG---RLLIQAASLIGWRAAIP-----RNAPGRCGCCQCAIGCPS 373
QY 261 -----EYVENHSAENLPHSIYNDKQKLFLLTNHRCTRLALTGG 301
DB 374 NAKFGVHLNALPOACAAGARIISWARVERILHR----- 406

QY 302 YEKIGAAEVRNLLATRNPSQLDSYIMAKVVVYLASGAIGNPQILYNSGFSGLQVTPRND 361
DB 407 -----AGRAYGVRRRRPDGTTLD--VLADAVVVAAGATETPCLLRSSGLGG----- 450
QY 362 SLIPNLGRYITEOPMAFCQIVLRQEFVDSVRDDPYGLPWKKEAVAQHIKAK---NPTDALP 418
DB 451 --HPRLGHNLALHPAT---MLAGLFDDDV-----FAW--RGVLQSAAVHEFHESDGV 496
QY 419 IPFRDPEPOVTTPTTEEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRW-----FGA 469
DB 497 IEATSTPPGM-----GSMVFFGYGAEL-----LRWLDRAPIATFGA 533
QY 470 TDEANNLIVFQNDYQDGYSMFQPT--FRY--RPSTASNVRR-----KMMADMCEVAS 519
DB 534 -----MVADRGVGTVRSVRGETVVRVDIAPGETAKLRVALQAIGRLLFAAGAVEVLT 585
QY 520 NGGYLPTSPQPMDPGL-----ALHLA-----CTTRIGFDKATTVADNNSLVWD 564
DB 586 G-----IPGAPPMRSLPELODVLRRANPRSLHLAAFHPTGTAAAGADEQLCPVDATGRLRG 641
QY 565 FANLYVAGNGTIRTGFGENPTLTSMCHAIKARSIIINTLKG 605
DB 642 VEGVWVADASILPSCPVEVNPQLSINAMALAVADQTVAKVVG 682

Search completed: September 3, 2002, 16:01:24
Job time: 333 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 15:49:55 ; Search time 54.14 Seconds
(without alignments)
1096.845 Million cell updates/sec

Title: US-09-856-327-2
Perfect score: 3284
Sequence: 1 MSLSTEOMLRDYPRSMQING.....IINTLKGTGDKNGTGEHRNL 618

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues 283138
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	200.5	6.1	573	2 E81385	probable oxidoredu
2	191.5	5.8	615	2 B38575	gluconate 2-dehydr
3	190.5	5.8	579	2 H87451	oxidoreductase, GM
4	187.5	5.7	619	2 AB2003	dehydrogenase chai
5	182	5.5	591	2 H83362	gluconate dehydrog
6	179.5	5.5	529	2 A84260	hypothetical prote
7	176	5.4	722	2 B75608	GMC oxidoreductase
8	171.5	5.2	562	2 JC7628	glucoside 3-dehydr
9	158.5	4.8	748	2 T10651	hypothetical prote
10	148.5	4.5	561	2 E98192	probable oxidoredu
11	148.5	4.5	561	2 AE3094	oxidoreductase Atu
12	138.5	4.2	527	2 E98202	dehydrogenase chai
13	138.5	4.2	527	2 AD3084	dehydrogenase Atu4
14	138.5	4.2	768	2 JC6564	cellobiose oxidase
15	138	4.2	494	2 AD3255	l-sorbose dehydrog
16	130.5	4.0	770	2 S60676	cellobiose oxidase
17	127.5	3.9	578	2 F70736	probable chod prot
18	124.5	3.8	578	2 D93400	probable oxidoredu
19	123	3.7	531	2 B83600	probable oxidoredu
20	121	3.7	1498	2 AF1082	B. subtilis Yuka p
21	117	3.6	736	2 A86171	hypothetical prote
22	116	3.5	502	2 T18562	hypothetical prote
23	116	3.5	514	2 B70917	probable zwf2 prot
24	116	3.5	599	2 T19711	hypothetical prote
25	114	3.5	1012	2 T00958	hypothetical prote
26	113.5	3.5	678	2 T05821	hypothetical prote
27	111.5	3.4	1042	2 T16169	hypothetical prote
28	110.5	3.4	589	2 T50698	probable mandelonl
29	110.5	3.4	5170	2 T15348	hypothetical prote

30	109.5	3.3	1125	2 T19193	hypothetical prote
31	108	3.3	1498	2 AG1439	B. subtilis Yuka p
32	107.5	3.3	557	2 F83181	probable GMC-type
33	107.5	3.3	567	2 AC0143	choline dehydrogen
34	107.5	3.3	964	2 S48404	probable membrane
35	107	3.3	1018	2 T40253	hypothetical prote
36	106.5	3.2	775	2 B72074	hypothetical prote
37	106.5	3.2	775	2 C81594	hypothetical prote
38	106.5	3.2	775	2 D86549	hypothetical prote
39	106.5	3.2	832	2 H84848	phospholipase D li
40	105.5	3.2	1611	2 T38236	hypothetical prote
41	105	3.2	499	2 T45749	hypothetical prote
42	104	3.2	509	2 D87452	oxidoreductase, GM
43	104	3.2	585	2 S72824	cholesterol oxidas
44	104	3.2	986	2 T33135	hypothetical prote
45	104	3.2	1917	2 C88728	protein C48A7.1 li

ALIGNMENTS

RESULT 1

E81385

Probable oxidoreductase chain Cj0415 [imported] - Campylobacter jejuni (strain NCTC 1

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000

C:Accession: E81385

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals

A:Reference number: A81250; MDID:20150912

A:Accession: E81385

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-573 <PAR>

A:Cross-references: GB:AL111168; GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB74251.1; PID:g696

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:

A:Gene: Cj0415

Query Match 6.1%; Score 200.5; DB 2; Length 573;

Best Local Similarity 20.6%; Pred. No. 3.5e-07;

Matches 135; Conservative 74; Mismatches 270; Indels 175; Gaps 26;

QY 36 VDFVIAGSGPIGATYAKLCVEAGLRVVMVEIG---AADSFYAVNAEEGTAVPYVPGVHKH 92

DB 8 VDVVTYAGWTGIVAAELTKAGLNLVSLERGHMQSTENFNIHDEWRYGINY----- 60

QY 93 NEIEFQKIDIDRFVNVIKGALQQVS---VPVRNQNVPTLDPGANSAPPSSAISNGKNPHQ 149

DB 61 -----GLMQDCSKDTVTFRH-----DPSGLALP----- 83

QY 150 REFENISAEAVTRGVGGMSTHTWCSTPRHPPMESLPGIGRP-----KLSNDPAEDD----- 201

DB 84 --YRKMGSFLLGNNGVAGVHNGWTFREFMPYDFEITLSKQRYGNKLGNDYTLQDMGVT 141

QY 202 -KENNELYSEARLIGTSYKEF-----DESIRHTLVLRSLQDAYKDRQRI 245

DB 142 YKDMEPYIDRFKTCGVSQEPNPLAEKMGAFRSPYPQEPLENTKMLKRFESAASNSLNH 201

QY 246 FRPLPLA-----CHRLKNAPEYVEVHSAENLPHSIYNDKQKK 283

DB 202 TYRLPASNKGKTYNPDGQDLAPCQVAYCERFG-----CEYGAASPLNTVTPKAMSTG 256

QY 284 LFTLLNHRCTRLLATGGYEKKIGAAAEVRNLLATRNPSOLDSYIM-AKYVVLASGAIGN 342

DB 257 KYTIRTVSNVTQIL-----KKDGKVTGVKVFVDTRT-----MKEYIQPADIVTLTSYMFNN 306

QY 343 PQILYNSGSGSLQVTP-----RNDSLIPNLG--RYITEQPMACQIVLROEFVDSVR 392

DB 307 AKLLAVSNI-GEQYDPKTKGTGLGRNYCYQNMNGTTAFDEQPNTF-----MGSGALGTTTS 361

QY 393 DDYGV--LPWKE-----AVAQHIANKPTDALPIPRDPPEQVTPPTTEHPWHTQIH 443
D 362 DDFNGDNFDSKEFLHGAMISVQLGTRPIQSADLPAGAP-----TWGAER- 408
QY 444 RDAFVG-----AVGPEVDSRVIVDLRFEGATDPANLLVFQNDVQDGYSPQPQTFYR 498
D 409 KALANYFTRATVGG-----GASLPHKNYLSLDPYTKDAFGMPLRLTYN 456
QY 499 PGTASNVARKMMAD-MCEVASNLGGLYPTSPQFMD-----PGLALHLAGTTRIGEDKA 552
D 457 -FTDQDRALHKKITDKTAEVARKMGVKSIGKAYILKDYVVPYOSTWNTGTTMGADRE 515
QY 553 TTVDANNLSLWDFANLYVAGNCTIRTFGENPTLFSMCHAKSARSIINTLKG 606
D 516 TSVWNTYLQHWADNLVYVVGAGNFQHNSGYNPTDVTGALAYCARCAGILKYHKS 569
RESULT 2
B38575
glucuronate 2-dehydrogenase (FAD) (EC 1.1.1.99.3) catalytic chain precursor [validated] - EMBL
C:Species: Erwinia cyripedii
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 08-Dec-2000
C:Accession: B38575; T52554
R:fum, D.Y.; Lee, Y.P.; Pan, J.G.
J. Bacteriol. 179, 6566-6572, 1997
A:Title: Cloning and expression of a gene cluster encoding three subunits of membrane-bound
A:Reference number: A38575; MUID:98012950
A:Accession: B38575
A:Molecule type: DNA
A:Residues: 1-615 <YUM>
A:Cross-references: GB:U97665; NID:g2584859; PIDN:AAC45885.1; PID:g2584862
A:Experimental source: strain ATCC 29267
A:Note: part of this sequence, including the amino end of the mature protein, was determined
A:Accession: T52554
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-615 <YU>
A:Cross-references: EMBL:U97665; PIDN:AAC45885.1
A:Experimental source: ATCC29267
C:Genetics:
A:Start codon: GTG
C:Complex: heterotrimer, consisting out of cytochrome c chain (T52555), catalytic chain
C:Function:
A:Description: EC 1.1.99.3 [validated, MUID:98012950]
C:Keywords: FAD; oxidoreductase
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-615/Product: gluconate 2-dehydrogenase catalytic chain #status experimental <MAT>

Query Match 5.8%; Score 191.5; DB 2; Length 615;
Best Local Similarity 21.1%; Pred. No. 1.9e-06;
Matches 134; Conservative 83; Mismatches 256; Indels 163; Gaps 32;

QY 36 VDVFAGSGPGATYAKLCVAGLRVWVEIGA-----ADSFYAVNAEGTAVVPYGYH 90
D 29 VDADVVGFGWAGAIMAKELTEAGLNVALERGPHRDTYPDGAYPOSIDELT-----YN 81
QY 91 KKNEIEFQKIDRFVNVIKGALQVSVPRNVQNTPLDPAWSAPPGSSAISNGKNPHQR 150
D 82 INKKL-FQ-DLSKSVTIRHDSQATVPYR----- 109
QY 151 EPENLASAVTRGVGMSTHTWCTSPRIHPHMESLPG-----IGRPKLSNDPAEDD--KE 203
D 110 ---QLAAFLPGTGTGAGLHSGVHFVRDVPVELNLSRHYEARYGKFIPEGWTIQDFGVS 166
QY 204 WNEL---YSEARLLGTSTKFEDESRITVLRSIQDAY--KDRQRIERPPLP-----LAC 253
D 167 VNELEPFDFQAEKVFGTSGSAW--TIKGMICKKEGKGNFYAPDRSSDF-PLPAQRRTYSA 223
QY 254 HRLKNAPVEVWH-----SAENLFHSIYND----- 279
D 224 QLFAQAESVGHYPDMPSANTSGPYTNTYGAQMGPFCNGYCSGYACMYSKASPNVNI 283

QY 280 ----KQKFLTLLTNHRCRTRALALGGYEKKIGA-----AEVRNLLATRNPSO--LDSYIM 329
D 284 LPALQEQEFELRNNAVYLRVNLTGDKKRCATCVTVLDQGREVV--QPADVLILSAFQF 340
QY 330 AKVYVLAGATGNP--QILYNSGFGSLQVTPRNDLSLPLN-GRYTEQPM---AFQIQLVR 384
D 341 HNVHLLMSLGGQPNPITNEGVVGRNFAYQNIISTLKALFDKNTTNPFIGAGGAGVAVD 400
QY 385 QEFVDSVRDDPYGL-----PWNKEAVAQHIANKPTDALPIPRDPPEQVTPPTTEHPWHT 440
D 401 DFNADNFQHPGTFVGGSPFW-----VNOAGTKPVSGLETPPKGTPN-----WGS 444
QY 441 QIHRDAFSYAGVPEVDSRVIVDLRFWFGATDPEANLLVFQNDVQDGYSP--QPTTFYR 498
D 445 QWK-----AAVADTYNHHISMDAH--CAHQSYRANLYLDLPNKYKNVTCQPLRLRTFDWQ 496
QY 499 PSTASNV-ARKMMADMCEV--ASN-----LGGYLTPSPQFMDPGL--ALHLAGTTRIGF 549
D 497 D---NDIRMAQFMVGKMRKITEAMNPKMIIG--AKGPGTHFDTTVYQTHMSGGAIMGE 551
QY 550 DKATTVDANNLSLWDFANLYVAGNCTIRTFGENPT 585
D 552 DPKTSAVNRVILQSWDVPNVFVPGASAFPOGLGYNPT 587
RESULT 3
H87451
oxidoreductase, GMC family CC1634 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: H87451
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: H87451
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-579 <STO>
A:Cross-references: GB:AE005673; NID:g13423038; PIDN:AAK23612.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC1634

Query Match 5.8%; Score 190.5; DB 2; Length 579;
Best Local Similarity 20.2%; Pred. No. 2.1e-06;
Matches 136; Conservative 83; Mismatches 270; Indels 185; Gaps 28;

QY 18 INGOIPKNAIHETYGNDGVDVFIAGSGPIGATYAKLCVAGLRVWVVEIGAADSFYAVNA 77
D 4 LNGRRRK---NTY-----DAIVVSGITGGIAAKELTEKGLKVLVL----- 42
QY 78 BEGTAVVPYGYHKKNETEFQKIDRFVNVIKGALQVSVPRNVQNTPLDPAWSAP 135
D 43 ERGPMVRHLEDY-----PTAMLDPMQSKYP 67
QY 136 PGSSA-----ISNGKNPHQRE--FENLSABAVTRG--VG 165
D 68 QGKLPEALNAHYKVQRTGYTMTQTOHFFVRDDEHPYTEENRFDWT-----RGYHVG 121
QY 166 GMSTHTWCTSPRIHPHMESLPGIGRPKLSND---PAEDDKWENLYSARLIGTSTKEF 222
D 122 GRSLTWGRQSTR-HSPID-FEANAREGIAVDWPIRYEDLAPW---YEHRFVIGVSGQA- 175
QY 223 DESIRH-----TLVLSLQDAYKDRQRI-FRPLPLACHRLKNAPYVSWHSA----- 268
D 176 -EGLPHFPDGHVQPPMELNCVEKAFKARSEARFPERRVTICRTAHLTDPTTEQALGRTK 234
QY 269 ---ENL-----FHSIYNDKQKKLFTLLTNHRCRTRALALGG-----YKKKIGAAEVR 312

Db 203 RKAABELGLHPYDLPAANASGPTWNPYGVQMGPCNFCGFCGACGYACMYKSKAPNL--NLL 260
QY 277 NDDKQKFLTLTNHRCRRLALTGGEKKIGA----AEVRNLLATRNPSQLDSYIMARV 332
Db 261 PALRQTPPLFELRANCNVLKVNLDSDGRQATGYVYVDAQGREIVQP-----AKL 308
QY 333 YVLASGATGNPOILYNSGSLQVTPRNDLSLIPNIGRITTEOPMAFCQIVLRQEVDSYR 392
Db 309 VIISAFQFHNVRLLLSGT-GRPYDPTGEGV--VCKNFAYONMA---TIKAFDFKDVH 361
QY 393 DDPY-----CLPWKKEAVAQHIAKNPTDALPIPFROPE 425
Db 362 TNPFGTGGGGVAVDDFNADNDHGPLGVGGSPMW----VNOAGSKPIGLGLVPPGTPS 417
QY 426 PQTTPFTTEHHPWT---QIHRDAFSYGAVGPEVDSRVIVDLRFWFGATDPEANLLVFQN 482
Db 418 -----WGSQWQAVKDAYTH-----TVSMDAHSNWTYRDNYLDLDP 454
QY 483 DVODGYSMP--OPTTRY-----RPTASNVRARKMMADMCEVA-----SNLGGYLPSTPP 530
Db 455 TYKDAYGQPLLRMTFDWKDNEIRMSRYVTEHMRKTAEAAMPKAISSVKNFGDHENTRVY 514
QY 531 QPMDPGLALHLAGTTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRTGFGENPTLTSWC 590
Db 515 Q-----TTHLLGGAIGSDPKTSVLNRYLQSDVHVNFVVGASAFPOGTGYNPTGLVAA 568
QY 591 HAIKSARSI 599
Db 569 LAYWSAKAI 577

RESULT 6
A84260
hypothetical protein Vng1035c [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: A84260
R:Ng, W.V.; Kennedy, S.P.; Mahalras, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: A84260
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-529 <STO>
A:Cross-references: GB:AE004437; NID:g10580587; PIDN:AAG19445.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG1035C

Query Match 5.5%; Score 179.5; DB 2; Length 529;
Best Local Similarity 22.1%; Pred. No. 1.3e-05;
Matches 143; Conservative 53; Mismatches 236; Indels 215; Gaps 28;

QY 37 DVFTAGSGPIGATYAKLCEAGLRVYVVEIGAADSFYAVNAEEGTAVPYVPGYHKKNEIE 96
Db 10 DVCVIGAGPAGGLADRLASDGHVSVVLEAG--PRFDAADR----- 50

QY 97 FOKDIDRVNVIKALQOVSVVVRNQNVPTLDPGWS--APPGSSAISNGKNPHOREFNL 155
Db 51 -----RMERSIRPAHGPAVY-----WEMGGPRDAYSTGD-----RYVPL 85

QY 156 SAEAVTRGVGMSTHTWCTSTPRIHP-----MESLPGIGRPKLSNDPAEDDKENNELYSAE 212
Db 86 NAARY-KVGSGSTLHWQGNVRLHQDPRLASATGVGA-----DWPPDYDFLKPYYAAAE 139

QY 213 RL1---GTSTKEFDESIRHTLVLSLODAYKDRQRIPLPLACHRLKNAPEYVEVWSAE 269
Db 140 SALGVSGASDNPFPAPPREQHPQAFPPSYSD--SLFAD-----ACESLGIAI----- 185

QY 270 NLFHSIYNDKQKKLFTLLTNHRCRRLALTG-----GYE-----K 304
Db 186 ---HSPVN-----AKLSAGRETRRACVGYGTQCPVPCPGAKYDATVHVDRATDAGAR 234
QY 305 KIGAAEVRNL-----LATRNPSSQLDSYIMAKYVYVVLASGAIGNPOIL----- 346
Db 235 VIDEAPVORLEHDAAGDRVTGAVYATPDGTTTHRS---ATEFVLAAGGIETPRLLLSDS 291
QY 347 --YNSGF---SGLQVTPRNDLSLIPNLGRITTEOPMAFCQIVLRQEVDSVRDPYGLPMW 401
Db 292 DRYPDGLANSGL-----VGRY-----FMD----- 311
QY 402 KEAVAQHIAKNPTDALPIFRDPEQVTPFTTEHPHWHTQIHRDAFSYGAV-----G 453
Db 312 -----HLFAGAGGTLDPETRQNHVGFNT--TESHQYV---DRPDGSRGAIKLEFLNYAG 360
QY 454 PEVDSRVIVDLRFWGA-----TPEANNLLVFONDVODGVSMPOQ 493
Db 361 FSPAEMALSGDDWDGAMCDRIDASGTHIAGVGLVEQQQPRPENRVRVLRHPERTDVHGNVP 420
QY 494 TFRYRPTASNVRARKMMADMCEVASNLGGYLP-TSPQFMDPGLALHLAGTTRIGFDKA 552
Db 421 DVVNSLSAYERITIERANEIQREILTELGADIEWTVGPE--DTGPAFHHMGTTRMGTDPA 478
QY 553 TTVDANNSLVWDFANLYVAGNGTIRTGFGENPTLTSMCHAIKSARSI 599
Db 479 ESVVDPRLRTHDLSNLSVASSVFPPTAGAMNPTLTIAALAKAADI 525

RESULT 7
B75608
GMC oxidoreductase - Deinococcus radiodurans (strain RI)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: B75608
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans RI.
A:Reference number: A75250; MUID:20036896
A:Accession: B75608
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-722 <WHI>
A:Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIDN:AAF12230.1; PID:g646
A:Experimental source: strain RI
C:Genetics:
A:Gene: DRA0127
A:Map position: 2

Query Match 5.4%; Score 176; DB 2; Length 722;
Best Local Similarity 20.9%; Pred. No. 3.8e-05;
Matches 139; Conservative 77; Mismatches 239; Indels 210; Gaps 33;

QY 14 RSMQINGQIPKNAIHTETGNDG---VDVFTAGSGPIGATYAKLCEAGLRVYVVEIGA- 68
Db 186 RNSAITPVVPO-----DGEVLEADAVVVGSGGGVIAARLAQAGKRVVLEAGGH 236

QY 69 --ABSFYAVNAEEGTAVPYVPGYHKKNEIEFQKIDIDRVNVIKALQOVSVVVRNQNVPT 126
Db 237 YHEAHFOGRELAAQTLYYRGYH-----PTADGNV-T 268

QY 127 LDPGAWSAPOGSSAISNGKNPHOREFNLSAEAVTRGVGMSTHTWCTSTPRIHPMESLP 186
Db 269 LVAGA-----NLG-----GGSTVWNSV-----PPRDDI- 293

QY 187 GIGRPKLSN-----DPAEDDKENNELYSAEARLIGTSTKEFDESIRHTLVLSLQDA 238
Db 294 ---RQRWASEHGLSDVADPGVD---RHIDAVLERM-GVSEQCSHDHNGPHQ---RLVEGA 342

QY 239 YKDRQIRFRPLPLACHRLKNAPEYVEVWSAENLFHSIYNDD---KQKKLFT----- 286

```
Db 343 DKLGTYTFVKA-----ALNLSPEH---YDADKAGHAGFGDQTGAQGTGTLNFKDAFEAG 393
QY 287 --LTLNRCTRLAUTGGEKIGAEVNRNLLATRNPSQLDSYIMAKVYVILASGAINPQ 344
Db 394 ARILVGTARQVLYEDG-----RAAGVSANTMGDETRQIT--VRAPQVAVVACALETPA 446
QY 345 ILYNSGSGGLQVTRPNDSLIPNLGRYITEQPMACQIVLRQEFVDSVRDDPYG---LPWW 401
Db 447 LLLRSGIGG-----PAAGRYLRLLHPAGLVAGI-----YGEDQRAWW 482
QY 402 KEAVAQIAKNPTDALPIPRDPE-----PQVIT---PFT--EEH--PWITQI 442
Db 483 GP-----PQSGILKQFADHENGHGFIEGVQYGPALMASGLPWTGGEAHRDLMSKF 533
QY 443 HRDAFSYGAVGPEVDSRVIVD-----LRWFGATDP--EANNLLVFONDVQDCYSM-----P 491
Db 534 HRMATFVSIQVDRGHGQVTVDEGNNAVHTYALTDLDARN---FRRGVTSIRLHEAAGA 590
QY 492 OPTFRYRPSTASNVARKMADMCVEAS---NLGGYLTPSPQPMDFGLALHLAGTTRIG 548
Db 591 EEIVALAPGAPVAPWRRGDLEAFIQVAQVPLGAGGQTVFS-----AHQMGSAWMG 640
QY 549 FDKATTVADNNSLVWDFANLYVAGNGIRTFGFGENPILTSMCHAIKSARSINTLKGTD 608
Db 641 SDPQTSVADPDQLHDVPGVWIGDTSAPFTCSGVNPMVSCMALASRTAEKLLAAMEGADG 700
QY 609 GKNTG 613
Db 701 GTSGS 705

RESULT 8
JC7628
glucoside 3-dehydrogenase (EC 1.1.99.13) - Halomonas sp. alpha-15
C;Species: Halomonas sp. alpha-15
C;Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001
C;Accession: JC7628; PC7123
R;Kojima, K.; Tsugawa, W.; Sode, K.
Biochem. Biophys. Res. Commun. 282, 21-27, 2001
A;Title: Cloning and expression of glucose 3-dehydrogenase from Halomonas sp. alpha-15
A;Reference number: JC7628; MUID:21164693; PMID:11263965.
A;Accession: JC7628
A;Molecule type: DNA
A;Residues: 1-562 <KOJ>
A;Accession: PC7123
A;Molecule type: protein
A;Residues: 2-11 <KO2>
C;Comment: This enzyme, as an oligomeric enzyme composed of catalytic and electron trans
and disaccharides.
C;Genetics:
A;Gene: g3dh
C;Keywords: oxidoreductase
```

```
Query Match 5.2%; Score 171.5; DB 2; Length 562;
Best Local Similarity 19.9%; Pred. No. 5.7e-05;
Matches 133; Conservative 85; Mismatches 236; Indels 215; Gaps 32;

QY 37 DVFTAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFTAVNAEEGTAVYPVGYHK----- 91
Db 7 DAIVVSGISGGWAAKELTEKGLVLL-----ERGRNIEHVKDYHNADKEA 53
QY 92 -----KNEIEFKQIDIRFVNVKCALQOVSVFVRNQNVPTLDP--GAWSAPPGSSAISNG 144
Db 54 WDPYRNE--PQEMIAKY-----PVLKRYPLNEATLGMWA-----DEQ 91
QY 145 KNPH--QREFENLSAEAVTRG--YGMSTHTWCTSTPRHPMESPGLGIRPKLSND---P 197
Db 92 ANPYEEKREFDWF-----RGYHVGRSLLWGRQSYRLSP--MDFEANREGIAIDWP 143
QY 198 AEDDKENNELYSAEERLIG--TSKEFDESHTVLVLSLQDAYKDRQIRFPLPLACHRL 256
```

```
Db 144 YEDLAPW---YDYVERFAGIAGTQE-----GLDILPDG-----EFLPPIPLNC--- 183
QY 257 KNAPEYVEHMSAEFLHSIYNDDKQKFLTLTLNHRCTRLALTGGEKIGKGAABVRNLLA 316
Db 184 -----VEEDAAXI-----KEAF-----GGQRLHLSRVANITQ 212
QY 317 TRNPSSQLDSYIMAKVYVILASGAINGPOILYNSGSGGLQVTRPNDSLIPNLGRYITEQPM 376
Db 213 PKPQNRVNCQYRNKCW-----LQCP---YGAVFSTQSATLPAAVATGNL---TLRPF 259
QY 377 AFQCIVL---RQ-----EPVDS-----VRDDPYGLPW-----WKEA 404
Db 260 SIVSQVLYKDRQRARGVEVIDAETHVEHYTADVIFLNASTFTNTWILMNSATDVMEGG 319
QY 405 VAQIAKNPTDALPIPR-----DPEQVTPPTFEHPWHQIHR-----DAF 447
Db 320 LGSSSGELGHNVDHFRCGASGEVEGLDKYYFCRRPAGFYIIFRNVGDEQSYVRGF 379
QY 448 SY-GAVGPEVDSRVIVDLR-----W-----FGATDPEANNLLVFONDVQD 486
Db 380 GYOGAASREGWDREIAELNIGADLKQALTQPGWTTIGTGFCEMLPDHNRISLDHSVRD 439
QY 487 GYSMPQPTFRYRPSTASNVARKMADMC-----VASNLGGLYLPSPQPMDFGL 537
Db 440 KWGLPVLISIDVELKQNERDMRDMQVQDAVDLLEAAGVKNKGVGDYA-----PGM 490
QY 538 ALHLAGTTRIGDFKATTVADNNSLVWDFANLYVAGNGIRTFGFGENPILTSMCHAIKSAR 597
Db 491 GIHEMGTARMGDRPKTSVLNSHNQVNDAPNVFVDGACMTSSSCVNPSTLTNALTARAVD 550
QY 598 SIINTLKGK 606
Db 551 YAVEELKRG 559

RESULT 9
TI0651
hypothetical protein T5F17.20 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Sep-2000
C;Accession: TI0651
R;Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Ban
submitted to the Protein Sequence Database, June 1999
A;Reference number: Z16533
A;Accession: TI0651
A;Molecule type: DNA
A;Residues: 1-748 <BEV>
A;Cross-references: EMBL:AL049917; GSPDB:GN00062; ATSP:T5F17.20
A;Experimental source: cultivar Columbia; BAC clone T5F17
C;Genetics:
A;Gene: ATSP:T5F17.20
A;Map position: 4
A;Introns: 70/3
C;Superfamily: Arabidopsis thaliana hypothetical protein T5K18.160
```

```
Query Match 4.8%; Score 158.5; DB 2; Length 748;
Best Local Similarity 19.4%; Pred. No. 0.00089;
Matches 126; Conservative 90; Mismatches 215; Indels 217; Gaps 28;

QY 37 DVFTAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFTAVNAEEGTAVYPVGYHKNEIE 96
Db 239 DAVVVGSGSGGVAANLAKAGLVLEKG---NYFTAHDYSGLEVPMSMLEYKQ--- 292
QY 97 FQKIDIRFVNVKCALQOVSVFVRNQNVPTLDPGAWSAAPPGSSAISNGKNPHQREFNLS 156
Db 293 -----GLLTV-----DCK-----FMLA 306
QY 157 AEAVTRGVGMSTHTWCTSTPRHPMESPGLGIRPKLSNDPAEDDKENNELYSAEERLIG 216
Db 307 GSAVG---GGTAVNWSAS---IRTPDHVL-----QEW-----SEGSKIKF 340
QY 217 TSTKEF----DE-SIRHTLVLSLQDAYKQKQIRFPLPLACHRL-----KNAPE-- 261
```

```
Db 341 FSGQEQSADVEYTRIGVTRCVKHGFQ-----OVLKRGGERLGLQVESVPRNSPEDH 395
QY 262 -----YVEWHSAENLFHSIYNDKOKKLFLLTHNCRTRAL-----TGGYEKK-----I 306
Db 396 YCGLCGYGCRAKNGTDGTLVDAVENGAVILTGKAERFVLVDNTSSNERKKRCVGV 455
QY 307 GAAEVNLLATRNPSQSDSYIMAKVYVLASGAIGNPQILYNSGSLGQVTPRNDSLIPN 366
Db 456 FASSVGGKIGKK-----FIIEARVTVSSAGSLTTPPLMLSSGLKN-----FN 497
QY 367 LGR-----YITQPMAFQCOIVLROEFVDSR--DDPYGLPWWKEAVAQHIKN 412
Db 498 IGRNKLKHPVMTWGYFPKSEFSKMEYEGGIITSVHHMD-----TESGCKRAILEN 550
QY 413 PTDALPIPRDPPOVYTPTEHPHWTQIHRDAFSYGAAGVPEVDSRVIVDLRWFGATDP 472
Db 551 PLIG-PASVAGLSPWVS-----GFDLKERMIK-----YGT-- 580
QY 473 EANNLVFONDQDGYSM--POPTFRYRPSTASNVRAKMMADMCEVAS---NLGGY--- 524
Db 581 --AHLFALVRLDLSGSEVMENEVYRTTKKDRNLRAGLRQALRVSAAGAVEVGYRSD 638
QY 525 -----LPTSPQFMDPLGAL-----HLAGTRIGFDKATTVADNN 559
Db 639 GQMKKEAITKAMEEFLDEDAVGVGTKGEYWTYFSAHQSGRCMGVTAEGALDEN 698
QY 560 SLVMDPANLYVAGNGTIRTGFGENPILTSMCHAIKSARSIIINTLKGGT 607
Db 699 GESWEAEGLFVCDGSLPSAVGNVPMITIQSTAYCISSKIVDSLQNK 746

RESULT 10
E98192
probable oxidoreductase chain CJ0415 [imported] - Agrobacterium tumefaciens (strain C58,
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: E98192
A:; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: E98192
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-561 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89063.1; PID:g15158859; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_980
A:Map position: linear chromosome

Query Match 4.5%; Score 148.5; DB 2; Length 561;
Best Local Similarity 19.1%; Pred. No. 0.0033;
Matches 135; Conservative 60; Mismatches 222; Indels 289; Gaps 29;

QY 37 DVFIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEEGTAVPVVGYHKKNEITE 96
Db 7 DAIVVSGISGGWAKELTQKGLKVLML-----ERGRNIEHTY----- 46
QY 97 FOKDIDRFVNVTKGALQOVSVPRVNONVPTLDPGAWSPGSSAISNGKNPHOREFENIS 156
Db 47 -----DKEAWDYPHRNRAQEMKAKYPVLSRDYL 78
QY 157 AEAVTRG-----VCGMSTHTCTSPRIHPHPMESLPGIGRPLK 193
Db 79 LEEATLGMWADQETPYVEEKRFDFRGVHGGRSLLMGRQTYRW--SQTDFEANA KGI 136
QY 194 SND---PAEDDKWENELSEARLIGTSTKEPDESIRHTLVLSLQDAYKDRQIRFRLP 250
Db 137 AVDWIRIEDVSPW---YDYVERFAGISGS-----REGLDILPDGE-FLPPIP 180
```

```
QY 251 LACHRLKNAPEVVEHNSAENLFHSIYNDKOKKLF---TLLNHRCTRRLALT----- 299
Db 181 L-----NEVEDQASRL-----KKAFTGRHLINSRCANITQELPDQERTP 221
QY 300 -----GGY-----EKKIGAAEVRNLLATR-----N 319
Db 222 COFRNKRCLGCGFGYFSTQASTLPAAVATGNLTLPFSIVKEILYDKDKKARGVEIID 281
QY 320 PSSQSDSYIMAKVYVLASGAIGNPQILYNSGF---SGLOVTPRNDSLIPNLGRYTE--- 373
Db 282 AETNLTYETADIIIFLNASTLNSVWLMNSATDVVEGGL-----GSSSGELGHVNDHH 335
QY 374 -----QPMAFQCOIVLROEFVDSR---DDP-----YGLP----- 399
Db 336 FRMGATGOVDFGDFEYFKGRRPAGF-----YIPRFRNTGDKRKYLRGFGYQGSASR 387
QY 400 --WWEAQAQHTAKNPTDALPIPRDPPEQVYTPTEHPHWTQIHRDAFSYGAAGVPEVD 457
Db 388 SWEREIAELNIGADYKDTL-----TEPGGW--TIGMTAF--GEMLYPHD 428
QY 458 SRVIVDLRWFGATDPEANLLVFQNDVQDGYSMPOPTFRYRPSTASNVRAKMMADMCE- 516
Db 429 NRVKLD-----HDKDKWGL-----PVLSMNVEKQNELDMRD 462
QY 517 -----VASNLGGLYPTSPQFMDPLGALHLAGTTRIGFDKATTVADNNS 560
Db 463 MYNDAVEMFEAVGIKNVKPSRGTYA-----PGMGIHENGATARMGRDPTSVLNGNN 513
QY 561 LYWDFANLYVAGNGTIRTGFGENPILTSMCHAIKSARSIIINTLKGG 606
Db 514 QVMDAPNVFVTDGACMTSASCYNPSSLTYNALTARAAEFAVSRKKG 559

RESULT 11
AE3094
oxidoreductase Atu4377 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AE3094
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; MCCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AE2577; PMID:11743193
A:Accession: AE3094
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-561 <KUR>
A:Cross-references: GB:AE008689; PIDN:AAL45171.1; PID:g17742847; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4377
A:Map position: linear chromosome

Query Match 4.5%; Score 148.5; DB 2; Length 561;
Best Local Similarity 19.1%; Pred. No. 0.0033;
Matches 135; Conservative 60; Mismatches 222; Indels 289; Gaps 29;

QY 37 DVFIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEEGTAVPVVGYHKKNEITE 96
Db 7 DAIVVSGISGGWAKELTQKGLKVLML-----ERGRNIEHTY----- 46
QY 97 FOKDIDRFVNVTKGALQOVSVPRVNONVPTLDPGAWSPGSSAISNGKNPHOREFENIS 156
Db 47 -----DKEAWDYPHRNRAQEMKAKYPVLSRDYL 78
QY 157 AEAVTRG-----VCGMSTHTCTSPRIHPHPMESLPGIGRPLK 193
Db 79 LEEATLGMWADQETPYVEEKRFDFRGVHGGRSLLMGRQTYRW--SQTDFEANA KGI 136
```


Db 109 VEH-----EDGVSPAWPVSYAEAEFFVYAEERLFGV-RGRAGDDPTEPPRSAPYMH 158
Qy 209 SEA---ERLIGTSTKDFESIRHTLVLSLQDAYKDRQIRPLPLA-----CHRLKN 258
Db 159 APIPEPIGRVAKGER-----LGLRPFH-----MPSAIDYGGGLCRRGCT 201
Qy 259 APEYVWEHSAENLFHSIYNDKQKKLF-----TLTNHRCRTRALTTGGYKKIGAA 309
Db 202 CDAFCVREDAGK-----DAETRLRLPALRHPNVSLTGARVRL-IAODGDKHIIVAV 252
Qy 310 EVRNLLATRNPSQLDSYIMAKVYVYVLAAGTGNPOILNYSFGLQVTPRNDLSLIPN--- 366
Db 253 EIER-----AGEITT-IEAPLFVLSAGAINSAIILERSA-----DEKPNGLA 294
Qy 367 -----LGRVITEQPM-----FCOIVLROEFVDSVRDDP----- 395
Db 295 NSSGVVGRYLMNHLSGLMLLPFTINDTRPKTMSLNDFFDGPGRGAAGNVQMLGNI 354
Qy 396 -----YGLPMWKEAVAQHIANKNPDTALPIPRDPEPQVTTPTTBEHPWHTQIHRDAFSY 449
Db 355 QCPMIRAAYPWMPRLANLLARHSVDFLVM-----SED----- 387
Qy 450 GAVGPEVDSRVLDLRFWGTGDPDPEANLLVFNQDVODGYSMQPQTFRYRPTASN-----V 505
Db 388 ---TPKYDSRV---KPMGKNGAE-----LIYRPGDREAHQRFV 419
Qy 506 RARKMADMCEVASNLGGYLTSPPOFMDPGLALHLAGTTRIGFDKATTVADNNSLVWDF 565
Db 420 RHMRSLLRKNFPVVLGHSGFIEAPS-----HQCCTVRMGDDPKAALNALCQTYDH 471
Qy 566 ANLYVAGNGTIRTGGENPTLTSCHAIK 594
Db 472 PNLYVVDAGFFPSSAALNPALTVAQAALR 500
RESULT 14
JC6564
cellobiose oxidase (EC 1.1.3.25) precursor - white-rot fungus (Trametes versicolor)
N:Alternate names: cellobiose dehydrogenase
C:Species: Trametes versicolor (white-rot fungus)
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 19-May-2000
C:Accession: JC6564
R:Dumonceaux, T.J.; Bartholomew, K.A.; Charles, T.C.; Moukha, S.M.; Archibald, F.S.
Gene 210, 211-219, 1998
A:Title: Cloning and sequencing of a gene encoding cellobiose dehydrogenase from Trametes
A:Reference number: JC6564; MUID:98248903
A:Accession: JC6564
A:Molecule type: mRNA
A:Residues: 1-768 <DOM>
A:Cross-references: GB:AF029668
A:Experimental source: 52J, ATCC 20869
C:Comment: This is a hemoflavoenzyme that oxidizes cellobiose and reduces iron(III) and
C:Genetics:
A:Gene: cdh
A:Introns: 17/1; 103/2; 140/1; 179/1; 226/3; 281/3; 306/1; 324/2; 534/3; 577/2; 593/3; 6
F:1-19/Domain: signal sequence #status predicted <SIG>
F:19/Domain: signal sequence #status predicted <SIG>
F:20-768/Product: cellobiose dehydrogenase #status predicted <MAT>
F:80,128/Binding site: heme iron (Met, His) (axial ligands) #status predicted
Query Match 4.2%; Score 138.5; DB 2; Length 768;
Best Local Similarity 19.2%; Pred. No. 0.032;
Matches 128; Conservative 83; Mismatches 228; Indels 229; Gaps 33;
Qy 37 DVFTAGSGPIGATYAKLCVEAGLRVWVVEIG-----AADSFYAVNAEEGTAVPY-VPG- 88
Db 231 DYIVVAGPGLVTDRLSEAGKGVLLERGGPSTAEETGGTYDATWAKSANLTKFDVPGL 290
Qy 89 ----YHKNETEFOKDIDRFVNLTKGALQQVSVPRNVNQTLDPGAWSAPPGSSAISNG 144
Db 291 FETLTDTNPFWMCKDNTFFAGCLLG-----GGTSV-NG 323

Qy 145 K---NPHOREFENLSAEAVTRGVGMSTHTWCSTPRIHPHMESLPGIGRKLSDPAEDD 201
Db 324 ALWYFNSKDFSTAS-----GWSSWSNHQFPTDKLQORLPS-----TDHPSADG 368
Qy 202 KENNELYSEA---ERLIGTSTKEFDESIRHTLVLSLQDAYKDRQIRF-----RP 248
Db 369 QRYLEQSAVTVQOQLLSGOGYSQI-----TINDNPDSDKHVGFSAFDELNGORA 417
Qy 249 LPLACHRLKNAPEYVWEHSAENLFHSIYNDKQKKLFTLLTNHRCRTRALTTGGYKKIGAA 308
Db 418 GPVA-----TYFETALARKNF---VYKDN---VLVTQVIRNGSTILGVRTN-DNTLGP 463
Qy 309 AEVRNLLATRNPSQLDSYIMAKVYVYVLAAGTGNPOILNYSFGLQVTPRNDLSLIP 365
Db 464 DGIVPL-----NPNGRV-----ILSGSGFTPRILFQSGIGTDMLOTVOSNAQAAA 510
Qy 366 NLGRYITEQPMAFQOIVLRQEFVSDRDPYGLPMWKEAVAQHIANKNPDTALPIPRDPE 425
Db 511 NL-----PP-----QSEWID-----LP-----VGQSVSDNPS----- 532
Qy 426 PQVITPTEHP-----WHTQIHRDAFSY-----GAVGPEVDSRVLDLRFW 466
Db 533 ---INLVFT---HPSIDAYDNWADWNSNPRADAQQYLOSRSVGLAGASPKLNF-----WRA 583
Qy 467 FGATDPEANLLVFQNDVQDGYSMQPQTFRYRPS-----TASN 504
Db 584 YGSGDQITR---YAQGVTRPGAASVNTSVAYNASEIFTITLYLSNGIOSRIGVDAALN 640
Qy 505 VRA-----RKMMADMCEVASNLGGYLTSPPOFM-----DPG---LA 538
Db 641 AKALVNPWLNTSVDKTVLLQALHDVTSTMKNVPGITMTIPDNTMTLEQYVAAVDPATMCS 700
Qy 539 LHLAGTTRIGFDKATTVADNNSLVWDFANLVAGNWTIRTGFGENPTLTSCHAIKSARS 598
Db 701 NHWVGAAMGVTSSSTAVVDENAKVFNTDLNLFIVDASIIPSILPIGNPQGVLM-SAAEQAVS 759
Qy 599 IINTILKGG 606
Db 760 RILALAGG 767
RESULT 15
AD3255
l-sorbose dehydrogenase (FAD) (EC 1.1.99.-) [imported] - Brucella melitensis (strain
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AD3255
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanov
; Mazur, M.; Goitsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A:Reference number: AD3252; PMID:11756688
A:Accession: AD3255
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-494 <CUR>
A:Cross-references: GB:AE008917; PIDN:AAL51207.1; PID:gl7981893; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0025
A:Map position: 1
C:Keywords: oxidoreductase
Query Match 4.2%; Score 138; DB 2; Length 494;
Best Local Similarity 20.4%; Pred. No. 0.017;
Matches 125; Conservative 67; Mismatches 240; Indels 182; Gaps 23;
Qy 37 DVFTAGSGPIGATYAKLCVEAGLRVWVVEIGAAADSFYAVNAEEGTAVPYVPGYHKKEIE 96
Db 6 DIVIIGSIGGATWAGLAAGLADILILEAGARLADRPENRPAI--FQGFPRKELW 63

```
QY 97 FOKDIDRFVNVKALQOVVPVRNQNVPTILDCAWSAPPSSAISNGK--NPHQREFEN 154
Db 64 YE-----TSGA-----PFNPGNYINVGNRSKFYGAVLIRYRREDFEE 100
QY 155 LSAAVTRGVGGSTHWCSTPRIHP---PMESLPGIGRPKLSNDPAE--DDKEWNELYS 209
Db 101 LA-----HLEGVSPAPFYSDELEPWYCKAEELFQV-RGELGDDPTPEPHSKRPYPYAAI 153
QY 210 EAERLICTSKEFDESIRHVLVRSLODAYKDRQIRFRPLPLACHRLKNAPEYVEWHSAE 269
Db 154 RDERPIADMRARLKKAGLHPASL-----PLGVDIERWL-AKARTPWDAHP 197
QY 270 NLFHSIYNDKQKLF-----LLTNHCTRL-----ALTGGYEKKIGAA 309
Db 198 NC-----DDGKMDAETCPAQAQYANVRLETSAQVRRLEAGPDGKAITAVHYVKNGEA 251
QY 310 EVRNLLATRNPSOLDSYIMAKVYVVLASGAIGNPQIILYNSGFSGLQVTPRNDLSLIPNLGR 369
Db 252 LV-----LRPKLVLSAGAVQSAALLLRSGLAN-----RSDQVGRNFMN 290
QY 370 YITEQPMAF-----CQIVLRQEF-----VDSVRDDPYGLPMWKEAVAQHIAKNPTDAL 417
Db 291 HNASAVIAFDPRYRNDSVYQKTFGFNDYILSDGAGGPPGLGNVOLLGRVSGAILKSNMRHV 350
QY 418 PIPFRPEPQVTPPTTEHPW-----HTQIHRDAFSYGAAGPEVDSRVIVD-----LR 465
Db 351 P-----EMLLNRIAGHT---IDFYAMSEDLFPSESRSVSDGDRIILH 389
QY 466 WFGATDPEANLLVFQNDVQGYSMPOPTFRYRPESTASNVRRKMMADMCEVASNLGGYL 525
Db 390 W-----VRSNWKHEHMLVLEKLSALRAVGF- 414
QY 526 PTSPQPMDFGLALHLAGTTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRTFGFGENPT 585
Db 415 PVVLSRAFDRTTPSHQCGTARIGNDPATAPLDPYCRAYDHPNLYVVDASFLPTSAAVNPA 474
QY 586 LTSMCHAIKSARSI 599
Db 475 LTAAQALRVADRI 488
```

Search completed: September 3, 2002, 15:58:45
Job time: 530 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 15:05:10 ; Search time 31.6 Seconds
(without alignments)
477.690 Million cell updates/sec

Title: US-09-856-327-2
Perfect score: 3284
Sequence: 1 MSLSTEQLRDPYRSMQING.....IINTLKGGTDGKNTGHRNL 618

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1232.5	37.5	623	1 US-08-734-925-2	Sequence 2, Appl1
2	1210	36.8	622	4 US-09-305-381-2	Sequence 2, Appl1
3	1091	33.2	566	4 US-09-023-731-1	Sequence 1, Appl1
4	192	5.8	51	4 US-09-023-731-12	Sequence 12, Appl1
5	191.5	5.8	615	4 US-09-297-937C-9	Sequence 9, Appl1
6	191.5	5.8	1276	4 US-09-297-937C-13	Sequence 13, Appl1
7	130.5	4.0	58	4 US-09-023-731-11	Sequence 11, Appl1
8	118.5	3.6	785	3 US-09-265-108-2	Sequence 2, Appl1
9	118.5	3.6	785	4 US-09-479-264-2	Sequence 2, Appl1
10	115	3.5	50	4 US-09-023-731-14	Sequence 14, Appl1
11	108	3.3	35	4 US-09-023-731-15	Sequence 15, Appl1
12	104.5	3.2	1876	2 US-08-609-049A-12	Sequence 12, Appl1
13	104.5	3.2	1876	4 US-09-170-996-12	Sequence 12, Appl1
14	102.5	3.1	61	4 US-09-023-731-6	Sequence 6, Appl1
15	100.5	3.1	806	3 US-08-549-515-5	Sequence 5, Appl1
16	100.5	3.1	806	3 US-08-549-515-11	Sequence 11, Appl1
17	100	3.0	664	4 US-08-485-284A-5	Sequence 5, Appl1
18	98.5	3.0	543	3 US-09-199-229-2	Sequence 2, Appl1
19	98.5	3.0	543	4 US-09-443-087-2	Sequence 2, Appl1
20	98.5	3.0	543	4 US-09-687-298-2	Sequence 2, Appl1
21	98	3.0	1257	1 US-08-049-783-2	Sequence 2, Appl1
22	98	3.0	1257	1 US-08-158-232-6	Sequence 6, Appl1
23	98	3.0	1257	1 US-08-304-626-6	Sequence 6, Appl1
24	98	3.0	1257	1 US-08-316-301A-6	Sequence 6, Appl1
25	98	3.0	1257	2 US-08-611-928-6	Sequence 6, Appl1
26	98	3.0	1257	3 US-09-173-891-6	Sequence 6, Appl1
27	98	3.0	1257	4 US-09-076-137-6	Sequence 6, Appl1

28	98	3.0	1257	5 PCT-US92-03624-6	Sequence 6, Appl1
29	97	3.0	3170	2 US-07-642-734C-5	Sequence 5, Appl1
30	97	3.0	3170	3 US-08-439-009A-5	Sequence 5, Appl1
31	96.5	2.9	1876	4 US-08-609-049A-28	Sequence 28, Appl1
32	96.5	2.9	1876	4 US-09-170-996-28	Sequence 28, Appl1
33	95.5	2.9	382	1 US-08-186-833-2	Sequence 2, Appl1
34	93.5	2.8	903	3 US-08-804-439A-22	Sequence 22, Appl1
35	93.5	2.8	903	3 US-08-720-229-2	Sequence 22, Appl1
36	93.5	2.8	904	4 US-08-632-537-1	Sequence 1, Appl1
37	93.5	2.8	904	5 PCT-US96-05316-1	Sequence 1, Appl1
38	93.5	2.8	904	6 5244792-4	Patent No. 5244792
39	93	2.8	604	2 US-08-746-283-32	Sequence 32, Appl1
40	93	2.8	604	2 US-08-746-283A-30	Sequence 30, Appl1
41	93	2.8	605	1 US-08-333-802-2	Sequence 2, Appl1
42	93	2.8	4545	2 US-08-804-227C-14	Sequence 14, Appl1
43	92.5	2.8	4472	2 US-08-804-227C-2	Sequence 2, Appl1
44	92	2.8	844	1 US-07-731-157A-6	Sequence 6, Appl1
45	92	2.8	844	2 US-08-541-780-6	Sequence 6, Appl1

ALIGNMENTS

RESULT 1
US-08-734-925-2
; Sequence 2, Application US/08734925
; Patent No. 5712139
; GENERAL INFORMATION:
; APPLICANT: NISHIMURA, IKUKO
; APPLICANT: OKADA, KIMIHARU
; APPLICANT: MINAMITHARA, TOMOYUKI
; APPLICANT: KAWAI, GENSUHIRO
; APPLICANT: KOYAMA, YASUJI
; APPLICANT: SUZUKI, MASARU
; TITLE OF INVENTION: NOVEL PYRANOSE OXIDASE, PYRANOSE OXIDASE
; TITLE OF INVENTION: GENE, NOVEL RECOMBINANT DNA AND PROCESS FOR PRODUCING
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: P.C.
; ADDRESS: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/734,925
; FILING DATE: 22-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/568,428
; FILING DATE: 06-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 7126-002-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 623 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-734-925-2

```

; ORGANISM: Corioliulus Versicolor
US-09-305-381-2

Query Match          36.8%   Score 1210;   DB 4;   Length 622;
Best Local Similarity 44.68; Pred. No. 2.1e-116;
Matches 27; Conservative 71; Mismatches 197; Indels 72; Gaps 16;

Qy 37 DVFIAGSPIGATYAKLCVEAGLVVMEIGAADSFYVNAEAGTAVPYYPGVYHKKNEIE 96
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48 DVAIVGSPICGTARELVEAGFNVMEIGEIDSLKI-----GSHKKNTVE 95

Qy 97 FOKDIDRVNVIKALQOVSVVPVNVQNVPTLPDGAASAPPGSSAINGKNPHOREFNLS 156
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 96 YQKNIDFVNVIOQLMPVSVPVNTVMVDTLSPASWQA--STFFVRNGAPEQDPLRNLS 153

Qy 157 AEATRVGGGNTHTWCTSTPRIHPMESLPGIGRPKL-SNDPAEDDKENWELISEARLI 215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 GOATRVVGGGNTHTWCATPRFK-----LQRPVLVKNDSKADDAEDRLYKKAESYF 206

Qy 216 GTSTKEFDESIRHPLVLSLQDAYKDRQIRFRPLPLACHRLKNAPEVVEVHSAENLF--H 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 207 KTGTTQFAESIRHNLVLKLOEEYKG-VRDFQOIPLAATR--QSPTFVESSAHTVPDLE 263

Qy 274 SIYNDQKQKFLTLTHNRCTRRLALTGGYEKKIGAAEVRNLLATRNPSOLDSYIMAKVY 333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 264 NRPNKDAPKORNFLPFAVACTNVRDNRNANSEIVG-LDVRDLHGKSKIT-----IKAKVY 316

Qy 334 VLASGATGNPOLLYNSFGSLQVTPRND-----SLTPNLGRYITEQPMAFQIVLRQEF 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 317 ILTAGAVHNAQLLAASFGQL---GREDPAPKPLPSLLPYLGTHTTEQTLVFCOTVMSTEL 373

Qy 388 VDSVRDDP--YGLP-----WKEAVAQHTAKNPTDALPIPPRDE 425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 374 INSTVADMTYVGKCHGHPDYSVTYTPGNPNKHPDWNKEVKHMDHQEDPLPIPEDPE 433

Qy 426 PQVTPTEHPHWTQIHRDAFSYGAVGEVDSRVIVDLRMFGATDPEANNLLVFQNDVQ 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 434 PQVTTLQATHPHWTQIHRDAFSYGAVQQSIDSRLIVDRFFGRTEPKEENKLFWSDKIT 493

Qy 486 DGYSNPQPTFYRSTASNVPRARKMDMCEVANLGGYLPSPQPMDPGLAHLAGTT 545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 494 DAYLRLQPTDFR--PFGGREAEEDMTDMCVMSAKIGGFLPGSYPOPMEPGLVHLGGTH 551

Qy 546 RIGFDKAT--TVADNNLSLWDFANLYVAGNCTIRTGFGENPTLTSMCHAKLSARSIIINTL 603
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 552 RMGDFEADKCCVDTDSRVFGFKLFLGGCGNIFTAYAAIPTLTAMSIAIKSEYI---K 608

Qy 604 KGGTDGKNTGHEHN 617
    | || || || || || || || || || || || || || || || || || || ||
Db 609 KNFEPSPNPVKHN 622

RESULT 3
US-09-023-731-1
; Sequence 1, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akhiro; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE

```

Query Match	37.5%;	Score 1232.5;	DB 1;	Length 623;
Best Local Similarity	45.0%;	Pred. No. 9.8e-119;		
Matches 273;	Conservative 73;	Mismatches 176;	Indels 85;	Gaps 16;

Qy	37	DVFITAGSPICATYAKLCVEAGLRVWVVEICAAOSFYAVNAEEGTAVPYVPGYHKHKEIE	96
Db	48	DWIVGSGPICCTYARELVEAGYKVAMPDICEIDSGLKI-----GAHKNTVE	95
Qy	97	FOKIDIRVNVIKGALQOVSVVVRNQNVPTLDPGAWSAPOGGSSAISNGKNPHQREFENLS	156
Db	96	YOKNIDFENVIQOLMSVSPVWTLVIDLSPISWQA--SSFFVRNGSNEQDPLRLMS	153
Qy	157	AEAVTRVGGGSMTHWTCTSPRIHPMESLPGIGRPKLSNDPAE--DKKEWNELYSEAEERLI	215
Db	154	GOAVTRVGGGSMTHWTCTATPFREDQ-----RELLVKQDQADDAEDRLYTKAESYF	206
Qy	216	GTSKPEDESIRHTVILRSLODAYKDRQIRPRPLPLACHRLKNAPEVYEHSAENLF--H	273
Db	207	KTGTDQPKESIRHNLVLNKLAEYKG-QRDQIQIPLAATR--RSPTFEVWESSANTVFDLQ	263
Qy	274	SIYNDKOKKLTLLTNHRCIRLALTGGYEKKIGAAEEVRNLLATRNPSQLDSY-----	327
Db	264	NRPTDAPNERFNLPVACERV-----VRN-----TSNSEIYESLHIHDLI	304
Qy	328	-----IMAKYVVLASGAIGNPQILYNSGFSGL-QVTPRN--DSLIPNLGRYITEQPMAFC	379
Db	305	SGDRFEIKADVEFLTAGAVHNAQLLVNSGFGQLGRPDPAHPQLLPSIGSYITEQSLVFC	364
Qy	380	QIVLRQPEVDSVRDDPY--GLP-----WKEVAOAHIAKNPTDAL	417
Db	365	QTVAMSTELIDSVKSDMIIRGNPGDLGSVYTPGAETNKHDPWNNEKYNHMMQHQEDPL	424
Qy	418	PIPFDRPEQVYTPFTEHPHHTOIHRDAFSYGAAGPEVDSRVIVDLRWGCATDPEANNL	477
Db	425	PIPFDRPEQVYTLFQSPHPHHTOIHRDAFSYGAAGQVDSRLIVDRFFGRTPEKEENK	484
Qy	478	LVFONDYQDGYSMPQPTFRYR-PSSTASNVARKMMADMCEVASNLGGYLPSPQFMDPG	536
Db	485	LWFSDKITDYNMQPITDFREPAGRSKEAEDMMTDMCVMSAKIGGFLPGSLPQFMEPG	544
Qy	537	LALHLAGTTRIGFDKA--TTVADNNLSLWMDPANLYVAGNGTIRTFGGENPTLTSMCHAIK	594
Db	545	LVLHLGGTHRMGEDEQEDKCCVNTDSRVFGKFLFGLGGCGNIPTAYGANPPLTAMSLAIK	604
Qy	595	SARSIIN 601	
Db	605	SCXEYIKN 611	


```

RESULT 2
US-09-305-381-2
; Sequence 2, Application US/09305381
; Patent No. 6146865
; GENERAL INFORMATION:
; APPLICANT: Christensen, Soren
; APPLICANT: Lassen, Soren Flensid
; APPLICANT: Schneider, Palle
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; TITLE OF INVENTION: Having Pyranose Oxidase Activity
; FILE REFERENCE: 5571.200-US
; CURRENT APPLICATION NUMBER: US/09/305,381
; CURRENT FILING DATE: 1999-05-05
; EARLIER APPLICATION NUMBER: 60/088,724
; EARLIER FILING DATE: 1998-06-10
; EARLIER APPLICATION NUMBER: PA 1998 00774
; EARLIER FILING DATE: 1998-06-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 622
; TYPE: PRT

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-023-731-1

Query Match 33.2%; Score 1091; DB 4; Length 566;
Best Local Similarity 42.6%; Pred. No. 4e-104;
Matches 248; Conservative 84; Mismatches 184; Indels 66; Gaps 15;

QY 37 DVFTAGSGPIGATYAKLCVE--AGLRVVMVEIGAADSFYVNAEEGTAVYVPGYHKKE 94
Db 33 DVFTAGSGPISCTYARHIDTSTKYMAEISQDN-----PVI-GAHHKS 79

QY 95 IEFQKIDRVNVVNIKALQOVSPVRNQNPTLDPGAWSPGSSA-----ISNGKNPHOR 150
Db 80 IKQKIDKFVNIINGALQISPSYQPTLAVAAW-APPIDPAGQLVIMCHNPQE 138

QY 151 EFENLSAEATRGVGMSTHTWCTSTRIHPMESLPQIGRPKLSNDPAEDDKWENLYSE 210
Db 139 AGLNPGSAVTRTVGGMATHWTACPTPHD-----EERVNDPV-DKQFDFALLER 187

QY 211 AERLIGSTKEEFESIRHTLVLSLODAYKDRORIFRPLPLACHRLKNAPEYVHWSAEN 270
Db 188 AKTLNVHSDQYDSDIRHQIVKETLQOTL-DASRGVTTPLGVERRTDNPIYVTWTGAD- 245

QY 271 LFHSIYNDDRQKKLFTLLTNHRCRLALTGGEKKIGAEVRNLLATRNPSSQLDSYMA 330
Db 246 ---TVLGDVPKSPREALVTETRTKLIVSETNPQVVAALLRNL---NTSN--DELVA 296

QY 331 KYVTLASGAIGNPOILYNSGSGIQTVPNDLSLIPNLGRYITQPMAFQCIQLRQEFVDS 390
Db 297 KSEVIACGAVCTPQILWNS-----NIRPY-----ALGRYLSEQSMFTFCQILVLRGIVDA 345

QY 391 VRDDPYGLPWKEVAQHIANKPTDALPIPRDPEPQVTTPTTEHPHWHTOIHRDAFSYG 450
Db 346 IATDPR-----FAKVEAHKKHPDDVLPFPFHEPEPQVMIPYTSDFPWHVQVHRDAFSYG 401

QY 451 AVGPEVDSRVIVDLRWFEGATDPEANLLVF-----QNDVQDGYSMPQPTFRYPST 501
Db 402 DVGPKADPRVVLDLRFEGKSDIYEENRVTGPNPKLREWEAGYTDYMGMPQTFHVKRTN 461

QY 502 ASNVRAKKMADCEVASNIGLYLPTSPQPMQDGLALHLAGTRIGTFDKATTVADNNSL 561
Db 462 ADGDRQDRQMMNDMTNVANMLGGYLPQSPQPMAPGLVHLITGTTRIGTDQTSVADPTSK 521

QY 562 VWDFANLYVAGNGTIRTFGENPTLTSMCHAIKSARSIIINTL 603
Db 522 VHNFNLMWVGNGCIPDATAACNPTRTSVAYALKGAEEAVVYL 563

RESULT 4
US-09-023-731-12
; Sequence 12, Application US/09023731

; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 51
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-023-731-12

Query Match 5.8%; Score 192; DB 4; Length 51;
Best Local Similarity 62.7%; Pred. No. 1e-12;
Matches 32; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 405 VAQHIANKPTDALPIPRDPEPQVTTPTTEHPHWHTOIHRDAFSYGAVGPE 455
Db 1 VEAHKKHPDDVLPFPFHEPEPQVMIPYTSDFPWHVQVHRDAFSYGDVGR 51

RESULT 5
US-09-297-937C-9
; Sequence 9, Application US/09297937C
; Patent No. 6337199
; GENERAL INFORMATION:
; APPLICANT: YUN, Do Young
; APPLICANT: PAN, Jae Gu
; TITLE OF INVENTION: Membrane-Bound Gluconate Dehydrogenase, Gene Sequence
; TITLE OF INVENTION: Encoding the Same and Production of 2-keto-D-Gluconate
; TITLE OF INVENTION: Using Transformed Recombinant E. Coli
; FILE REFERENCE: P66159US0
; CURRENT APPLICATION NUMBER: US/09/297,937C
; CURRENT FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: PCT/KR98/00296
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: KR 97-48802
; PRIOR FILING DATE: 1997-09-25
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 615
; TYPE: PRT

; ORGANISM: Erwinia cyripriedii
US-09-297-937C-9

Query Match 5.8%; Score 191.5; DB 4; Length 615;
Best Local Similarity 21.1%; Pred. No. 9.5e-11;
Matches 134; Conservative 83; Mismatches 256; Indels 163; Gaps 32;
QY 36 VDVFIAGSPIGATYAKLCVEAGLRVVMVEIGA-----ADSFYAVNAEEGTAVPVPGVH 90
DB 29 VDAVVVGCGWAGAIMAKELTEAGLNVVALERGRPHRTYPDGAYPOSIDELT-----YN 81
QY 91 KNEITEFQKIDRFNVNIKALQOVSVVPRNVNPTLDPGAWSPGSSAISNGKNPHQR 150
DB 82 IRKKL-FQ-DLSKSVTIRHDASQTAVPYR----- 109
QY 151 EFENLSAEAVTRGVGSMTHWCTSPRIHPMESLPG-----IGRPKLSNDPAEDD--KE 203
DB 110 ---QLAAFLPGTGTGAGLHWSGVHFRVDPVELNLSHYEARYGKNFPEGMTIODFGVS 166
QY 204 WNEL---YSEARLIGTSTKEFDESIRHTLVLSLQDAY--KDRQIRFRLP-----LAC 253
DB 167 YNELEPFQDAEKVFGTSSAW--TIKGMIGKEKGGNFYAPDRSSDF-PLPAQRRTYSA 223
QY 254 HRLKNAPEYVEWH-----SAENLFHSIYND----- 279
DB 224 QLFAQAESVGHYPDMPSANTSGPYTNTYGAQMPCNFCGCGYACMYSKASPNVNI 283
QY 280 ----KQKLLTLLNHRCTRLALTGGYEKKICA-----AEVRNLLATRNPSQ--LDSYIM 329
DB 284 LPALRQEPKFEURNAYVLRNLTGDKKRATGVTVLDGGREV---QPADLVILSAFQF 340
QY 330 AKVYVLAGSAGTGNP-QILYNSGFSGLQVTPRNDLSLIPNL-GRYITEQPM---AFQCIIVLR 384
DB 341 HNVHMLLSGIGQPNITNEGVGRNFAYQNIISLKALFDKNTTNPFIGAGGAGVAVD 400
QY 395 QEFVDSVRDDPYGL-----PWWKEAVAQIAKNPTDALPIPRDPPEQVTPTEHPHWT 440
DB 401 DENADNFDHGPYGFVGGSPFW---VNOAGTKPVSGCLTPKGTN-----WGS 444
QY 441 QIHRDAFSYAGVPEVDSRVIVDLRFMGATDPEANLLVFQNDVDGYSM-PTFRYR 498
DB 445 QWK-----AAVDYTNHHISMDAH--GAHOSYRANYLDLDPYKNVYGOPLLRMTFDMQ 496
QY 499 PSTASNR-ARKMADMCVEV-ASN-----LGYLPTSPQFMDPGL--ALHLAGTTRIGF 549
DB 497 D---NDIRMAQPMVGMKMKITEAMNPKMIIGG--AKGPGTHFTTYYQTHMSGGAIMGE 551
QY 550 DKATTVADNNSLVDFANLYVAGNCTIRTFGENPT 585
DB 552 DPKTSAVNRYLQSDWDPVNVFVPGASAFQGLGYNPT 587

RESULT 6
US-09-297-937C-13
; Sequence 13, Application US/09297937C
; Patent No. 6337199
; GENERAL INFORMATION:
; APPLICANT: YUM, Do Young
; APPLICANT: PAN, Jae Gu
; TITLE OF INVENTION: Membrane-Bound Gluconate Dehydrogenase, Gene Sequence
; TITLE OF INVENTION: Encoding the Same and Production of 2-Keto-D-Gluconate
; TITLE OF INVENTION: Using Transformed Recombinant E. Coli
; FILE REFERENCE: P66159US0
; CURRENT APPLICATION NUMBER: US/09/297,937C
; CURRENT FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: PCT/KR98/00296
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: KR 97-48802
; PRIOR FILING DATE: 1997-09-25
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13

; LENGTH: 1276
; TYPE: PRT
; ORGANISM: Erwinia cyripriedii
US-09-297-937C-13

Query Match 5.8%; Score 191.5; DB 4; Length 1276;
Best Local Similarity 21.1%; Pred. No. 3.4e-10;
Matches 134; Conservative 83; Mismatches 256; Indels 163; Gaps 32;
QY 36 VDVFIAGSPIGATYAKLCVEAGLRVVMVEIGA-----ADSFYAVNAEEGTAVPVPGVH 90
DB 249 VDAVVVGCGWAGAIMAKELTEAGLNVVALERGRPHRTYPDGAYPOSIDELT-----YN 301
QY 91 KNEITEFQKIDRFNVNIKALQOVSVVPRNVNPTLDPGAWSPGSSAISNGKNPHQR 150
DB 302 IRKKL-FQ-DLSKSVTIRHDASQTAVPYR----- 329
QY 151 EFENLSAEAVTRGVGSMTHWCTSPRIHPMESLPG-----IGRPKLSNDPAEDD--KE 203
DB 330 ---QLAAFLPGTGTGAGLHWSGVHFRVDPVELNLSHYEARYGKNFPEGMTIODFGVS 386
QY 204 WNEL---YSEARLIGTSTKEFDESIRHTLVLSLQDAY--KDRQIRFRLP-----LAC 253
DB 387 YNELEPFQDAEKVFGTSSAW--TIKGMIGKEKGGNFYAPDRSSDF-PLPAQRRTYSA 443
QY 254 HRLKNAPEYVEWH-----SAENLFHSIYND----- 279
DB 444 QLFAQAESVGHYPDMPSANTSGPYTNTYGAQMPCNFCGCGYACMYSKASPNVNI 503
QY 280 ----KQKLLTLLNHRCTRLALTGGYEKKICA-----AEVRNLLATRNPSQ--LDSYIM 329
DB 504 LPALRQEPKFEURNAYVLRNLTGDKKRATGVTVLDGGREV---QPADLVILSAFQF 560
QY 330 AKVYVLAGSAGTGNP-QILYNSGFSGLQVTPRNDLSLIPNL-GRYITEQPM---AFQCIIVLR 384
DB 561 HNVHMLLSGIGQPNITNEGVGRNFAYQNIISLKALFDKNTTNPFIGAGGAGVAVD 620
QY 385 QEFVDSVRDDPYGL-----PWWKEAVAQIAKNPTDALPIPRDPPEQVTPTEHPHWT 440
DB 621 DENADNFDHGPYGFVGGSPFW---VNOAGTKPVSGCLTPKGTN-----WGS 664
QY 441 QIHRDAFSYAGVPEVDSRVIVDLRFMGATDPEANLLVFQNDVDGYSM-PTFRYR 498
DB 665 QWK-----AAVDYTNHHISMDAH--GAHOSYRANYLDLDPYKNVYGOPLLRMTFDMQ 716
QY 499 PSTASNR-ARKMADMCVEV-ASN-----LGYLPTSPQFMDPGL--ALHLAGTTRIGF 549
DB 717 D---NDIRMAQPMVGMKMKITEAMNPKMIIGG--AKGPGTHFTTYYQTHMSGGAIMGE 771
QY 550 DKATTVADNNSLVDFANLYVAGNCTIRTFGENPT 585
DB 772 DPKTSAVNRYLQSDWDPVNVFVPGASAFQGLGYNPT 807

RESULT 7
US-09-023-731-11
; Sequence 11, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-023-731-11

Query Match 4.0%; Score 130.5; DB 4; Length 58;
Best Local Similarity 44.4%; Pred. No. 3.1e-06;
Matches 28; Conservative 11; Mismatches 13; Indels 11; Gaps 2;

QY 333 YVLASGAIGNPQILYNSGFGSLQVTPRNDLSLIPNLGRYIEQPMAPCOIVLRQEFVDSVR 392
Db 2 FVIACGAVCTPQILWNS-----NIRPY-----ALGRYLSQSMTCQIVLKRGIYDAIA 50
QY 393 DDP 395
Db 51 TDP 53

RESULT 8
US-09-265-108-2
; Sequence 2, Application US/09265108
; Patent No. 6033891
; GENERAL INFORMATION:
; APPLICANT: Gollightly, Elizabeth
; APPLICANT: Brown, Kimberly
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; FILE REFERENCE: 5850.000-US
; CURRENT APPLICATION NUMBER: US/09/265,108
; EARLIER FILING DATE: 1999-03-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 785
; TYPE: PRT
; ORGANISM: Humicola
US-09-265-108-2

Query Match 3.6%; Score 118.5; DB 3; Length 785;
Best Local Similarity 19.5%; Pred. No. 0.0055;
Matches 126; Conservative 93; Mismatches 257; Indels 171; Gaps 30;

QY 26 AIHETYNDGVDFPIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYVNAEEGTAVP- 84
Db 242 APEETY-----DYIVVGAGAGGIPVADKLSEAGHKVLLIEKGP-----STGRWQGTMKPE 292
QY 85 YVPGYHKKNEIEFOKIDRFVNVIKGALQOVSVPRNQNVPTLDPGAWSPGSSAISNG 144
Db 293 WLEG-----TDLTRF--DVPGLCNQIWDSDAGIACITDQDAGCVLGGGTAVNAG 340

; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 58
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-023-731-11

Query Match 4.0%; Score 130.5; DB 4; Length 58;
Best Local Similarity 44.4%; Pred. No. 3.1e-06;
Matches 28; Conservative 11; Mismatches 13; Indels 11; Gaps 2;

QY 333 YVLASGAIGNPQILYNSGFGSLQVTPRNDLSLIPNLGRYIEQPMAPCOIVLRQEFVDSVR 392
Db 2 FVIACGAVCTPQILWNS-----NIRPY-----ALGRYLSQSMTCQIVLKRGIYDAIA 50
QY 393 DDP 395
Db 51 TDP 53

RESULT 9
US-09-479-264-2
; Sequence 2, Application US/09479264
; Patent No. 6280976
; GENERAL INFORMATION:
; APPLICANT: Elizabeth J. Gollightly
; APPLICANT: Kimberly M. Brown
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; FILE REFERENCE: 5850.200-US
; CURRENT APPLICATION NUMBER: US/09/479,264
; CURRENT FILING DATE: 2000-01-05
; EARLIER FILING DATE: 1999-03-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 785
; TYPE: PRT
; ORGANISM: Humicola
US-09-479-264-2

Query Match 3.6%; Score 118.5; DB 4; Length 785;
Best Local Similarity 19.5%; Pred. No. 0.0055;
Matches 126; Conservative 93; Mismatches 257; Indels 171; Gaps 30;

QY 26 AIHETYNDGVDFPIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYVNAEEGTAVP- 84
Db 242 APEETY-----DYIVVGAGAGGIPVADKLSEAGHKVLLIEKGP-----STGRWQGTMKPE 292
QY 85 YVPGYHKKNEIEFOKIDRFVNVIKGALQOVSVPRNQNVPTLDPGAWSPGSSAISNG 144
Db 293 WLEG-----TDLTRF--DVPGLCNQIWDSDAGIACITDQDAGCVLGGGTAVNAG 340

QY 145 K--NPHOREFENLSAEAVTRGVGCMSTHTWCSTPRIHPPMESLPFGIGRPKLSNDPAED-- 200
Db 341 LWMKPIDLWDENFPE-----GWHSQDLAAATERV---FERIPGTWHPSDGKLYRDEG 391
QY 201 -----DKWNELYSEAEIRLGTSTKEFDESIRHTLVLSLQDAYKDKQRIFRPLPL 251
Db 392 YKVLSSGLAESGWKEV-----VANEVNPENKRTFAHTHFAGCE-----RNGPL 436
QY 252 ACHRLKNAPEYVEWHSANLFLHSIYNDKOKKFLFTLTNHRCTRLALTGGYKKGAAEV 311
Db 437 ATY-----LVSADAREN-----PSLATNTAVRAVRTGG---KVTGVEL 472
QY 312 RNLLATRNPSQLDSYIMAKYVVLASGAIGNPQILYNSGFGSLQVTPRNDLSLIPNLGRYI 371
Db 473 ECL--TDGYSIGVILKNEGCVIFISAGAFSGAKLLFRSGIG----- 511
QY 372 TEQPMAPCOIVLR-----QEFVDSVRDDPYGLPMWKEAAVOAHIKN-PTDAL---PIPER 422
Db 512 ---PEDQLRVVASSKOGEDFIDE-KD-----WIKLPVGNLIDHLNWDLTHTPDVVFY 561
QY 423 DPEQVTTPTTEHPWHTQIHRDAFSYGA--VGP-----EVDSRVIVDL 464
Db 562 DFEAWTTPTEADKQLYLEQRSGLIAQAAPNIGPMMEQVTPSDGITRQFQWTVARVSGDS 621
QY 465 RWFQATDPEANLLVFONDVODGYS-----MPQPTFRYRPSTASNVNARKMMADM 514
Db 622 RFTNSSHAMTLSQYLRGVVSRGRATITQGLVTTVAEHPYLHNAGDKEAVIQGKLNIES 681
QY 515 CEVASNLGGVLPSTSP-----QFMDPGL-----ALHLAGTTRIGFDK---ATTVADN 558
Db 682 LNVIPNITWLP--PPGSTVEEYVDSLLVSASARRSNHNMGTAKLGTDGGRYGTSVVDL 739
QY 559 NSLWVDPFANLYVAGNGTIRTGFGGE-NPTLTSMCHAISARSIIITLK 604
Db 740 DTKVYGTDLNLFVW-DASIFFGCMSTGPNPSAMIVIAAEQAERILKLRK 785

RESULT 9
US-09-479-264-2
; Sequence 2, Application US/09479264
; Patent No. 6280976
; GENERAL INFORMATION:
; APPLICANT: Elizabeth J. Gollightly
; APPLICANT: Kimberly M. Brown
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
; FILE REFERENCE: 5850.200-US
; CURRENT APPLICATION NUMBER: US/09/479,264
; CURRENT FILING DATE: 2000-01-05
; EARLIER FILING DATE: 1999-03-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 785
; TYPE: PRT
; ORGANISM: Humicola
US-09-479-264-2

Query Match 3.6%; Score 118.5; DB 4; Length 785;
Best Local Similarity 19.5%; Pred. No. 0.0055;
Matches 126; Conservative 93; Mismatches 257; Indels 171; Gaps 30;

QY 26 AIHETYNDGVDFPIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYVNAEEGTAVP- 84
Db 242 APEETY-----DYIVVGAGAGGIPVADKLSEAGHKVLLIEKGP-----STGRWQGTMKPE 292
QY 85 YVPGYHKKNEIEFOKIDRFVNVIKGALQOVSVPRNQNVPTLDPGAWSPGSSAISNG 144
Db 293 WLEG-----TDLTRF--DVPGLCNQIWDSDAGIACITDQDAGCVLGGGTAVNAG 340

QY 145 K--NPHOREFENLSAEAVTRGVGCMSTHTWCSTPRIHPPMESLPFGIGRPKLSNDPAED-- 200
```

Db 341 LWWKPIDLDWENFPE-----GWSQDLAAATERY---FERIPGTWHPMSMDGKLYRDEG 391
QY 201 -----DKWNNELYSAEARLIGTSKEFDESIRHTLVLSLODAYKDRORIRPLPL 251
Db 392 YKVLSSGLAESGWKE-----VANEVPEKRTFAHTHEMFAGGE-----RNGPL 436
QY 252 ACHRLKNAPEYVEHSAENLFHSIYNDOKKLFLLTNHRCITRLALTGGEYKKGIGAAEV 311
Db 437 ATY-----LVSADAREN-----FSLWNTAVRAVRIGG---KVTGVEL 472
QY 312 RNLATRNPSOLDYSIMAKVYVLASGAIGNPQILYNSGFLGVTPTNRNDSLIPNLGRYI 371
Db 473 ECL--TDGYSIGIVKLNEGGSVIFGAGAFGSAKLLFRSGIG-----511
QY 372 TEQPMACQIVLR---OEFVDSVRDDPYGLPWKEAQAHIKAKN-PTDAL-----PIPR 422
Db 512 ----PEQURVAVASSKGDGFIDE-KD-----WIKLPVGYNLIDHLNLTDLILTHFDVVFY 561
QY 423 DPEQVTPPTFEHPWHQIHRDAFSYGA--VGP-----EVDNRVIVDL 464
Db 562 DFEAWTTPIEADKOLYLEQRSGILAQAAPNIGPMMEQVTPSDGITRQFQWTVARVEGDS 621
QY 465 RWFEGATDPEANLLVFQNDVQDYS-----MPOPTFRYRSTASNVRAKMMADM 514
Db 622 RFTNSHAMTLSQLGRGVYVSRGRATITQGLVTVAEHPYHLNAGDKAEAVIQIKNLIES 681
QY 515 CEVASNLGGYLTSP-----OFMDPGL-----ALHLAGTTRIGFDK-----ATTVDN 558
Db 682 LNVIRNITWVLP--PFGSTVEYVDSLLVSASARSNNHMGFAKIGTDDGRYGGTSVVDL 739
QY 559 NSLVWDFEANYVAGNGTIRTGEGE-NPILTSCHAIKASRSINTLK 604
Db 740 DTKVGTGDNLFVV-DASIFPGHSTGNPSAMIVIAEQAERILKLR 785

RESULT 10

US-09-023-731-14

; Sequence 14, Application US/09023731

; Patent No. 6291648

; GENERAL INFORMATION:

; APPLICANT: Kawamura, Yukio; Morita,

; APPLICANT: Akhihiro; Izumo, Koji.; Saka, Tomohide.

; TITLE OF INVENTION: ANTITUMOR PROTEIN AND

; TITLE OF INVENTION: GENE ENCODING SAME

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.

; STREET: 345 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: MICROSOFT WORD 97

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/023,731

; FILING DATE: 13-FEB-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 29275/1997

; FILING DATE: 13-FEB-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: KENNETH H. SONNENFELD

; REGISTRATION NUMBER: 33,285

; REFERENCE/DOCKET NUMBER: 3479-4000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 758-4800

; TELEFAX: (212) 751-6849

; TELEX: 421792

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 50

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-09-023-731-14

Query Match 3.5%; Score 115; DB 4; Length 50;

Best Local Similarity 53.5%; Pred. No. 9.7e-05;

Matches 23; Conservative 5; Mismatches 15; Indels 0; Gaps 0;

QY 484 VODGYSMQPTFRYRSTASNVRAKMMADMCEVASNLGGYLP 526

Db 8 VTDTYGMQPTFHVRTNADGDRDQRMNDMTNVANMLGGYLP 50

RESULT 11

US-09-023-731-15

; Sequence 15, Application US/09023731

; Patent No. 6291648

; GENERAL INFORMATION:

; APPLICANT: Kawamura, Yukio; Morita,

; APPLICANT: Akhihiro; Izumo, Koji.; Saka, Tomohide.

; TITLE OF INVENTION: ANTITUMOR PROTEIN AND

; TITLE OF INVENTION: GENE ENCODING SAME

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.

; STREET: 345 PARK AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10154

; COMPUTER READABLE FORM:

; MEDIUM TYPE: FLOPPY DISK

; COMPUTER: IBM PC COMPATIBLE

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: MICROSOFT WORD 97

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/023,731

; FILING DATE: 13-FEB-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 29275/1997

; FILING DATE: 13-FEB-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: KENNETH H. SONNENFELD

; REGISTRATION NUMBER: 33,285

; REFERENCE/DOCKET NUMBER: 3479-4000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 758-4800

; TELEFAX: (212) 751-6849

; TELEX: 421792

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 35

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-09-023-731-15

Query Match 3.3%; Score 108; DB 4; Length 35;

Best Local Similarity 66.7%; Pred. No. 0.00028;

Matches 22; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 528 SPPQPMDFGLALHLGAGTTRIGFDKATTVADNNS 560

Db 2 SYPQPMAGPLVLHITGTTTRIGTDDQTSVADPTS 34

RESULT 12

US-08-609-049A-12

; Sequence 12, Application US/08609049A
; Patent No. 5948664
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 5948664el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 29-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-063700US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1876 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-609-049A-12

Query Match 3.2%; Score 104.5; DB 2; Length 1876;
Best Local Similarity 22.0%; Pred. No. 0.72;
Matches 85; Conservative 53; Mismatches 170; Indels 79; Gaps 21;

QY 83 VPYPGVYHKKNEIEFQKIDRFVNVKIGALQVSVPRNQNVPDTPGA----WSAP--P 136
||| | : : : : : ||| | : : : : :
DB 184 VPYPAQAQQORPLN-SEELQRLYSM---PAQMAVVPVQPNAYMYPGAVVPTATPIVP 239
||| | : : : : : ||| | : : : : :
QY 137 GSSAISNGKNPHOREFENLSAEAVTRGVGMSTHTWCTPRIHP-PMESLPGIG--RPKL 193
||| | : : : : : ||| | : : : : :
DB 240 GSAAFMPQYPAQ-----GYGFGGAYTHMDLRRPQSPAPQOATPTTSHHSQP 288
||| | : : : : : ||| | : : : : :
QY 194 SNDAEDDKENLYSEAEER----LIGTSKEFEDESIRHTLVLSLQDA----YKDRQRI 245
||| | : : : : : ||| | : : : : :
DB 289 SNHSTSPAPANGVAFPARQVFTVGVSSTHTGNGSHSVPRRNDLIDLNHEDYSRV 348
||| | : : : : : ||| | : : : : :
QY 246 -----FRPLPLACHRLKNAPEYVHWSAENLHFSIYNDKQKKLFTLLTNHCRTRLALTG 300
||| | : : : : : ||| | : : : : :
DB 349 SVLEAFDPLNDNTGNTASDTSIYAEYDFDFLYSGD-----AATQYSDPMYEAVN 401
||| | : : : : : ||| | : : : : :
QY 301 GYEKKIGAAEVR-NLLATR-----NPSQLDSYIMA-----KVYVLASGAIGNP----- 343
||| | : : : : : ||| | : : : : :
DB 402 RWDKTATVSPNVGLIGWRQDFLSQSTSSSYQGVAPPEESLKLAEANGSGTISPPPLPP 461
||| | : : : : : ||| | : : : : :
QY 344 --QILYNSGFGSLQVLT-PRNDSLIPNLGRYITEOPMAFCOIVLRQ-----EFVDSVR 392
||| | : : : : : ||| | : : : : :
DB 462 RNOOCYESNOAMPVSRPPOSSVLTLD---SYTSSIP---ANVLDRRKTKTRLVELISDQR 516
||| | : : : : : ||| | : : : : :
QY 393 -DDPYGLPWKEAQAQIAKNPTDALP 418
||| | : : : : : ||| | : : : : :
DB 517 TDDPELLEFY-HMWKEVRYRPHDDAP 542
||| | : : : : : ||| | : : : : :

RESULT 13
US-09-170-996-12
; Sequence 12, Application US/09170996
; Patent No. 6291220
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 6291220el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 29-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-063700US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1876 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-170-996-12

Query Match 3.2%; Score 104.5; DB 4; Length 1876;
Best Local Similarity 22.0%; Pred. No. 0.72;
Matches 85; Conservative 53; Mismatches 170; Indels 79; Gaps 21;

QY 83 VPYPGVYHKKNEIEFQKIDRFVNVKIGALQVSVPRNQNVPDTPGA----WSAP--P 136
||| | : : : : : ||| | : : : : :
DB 184 VPYPAQAQQORPLN-SEELQRLYSM---PAQMAVVPVQPNAYMYPGAVVPTATPIVP 239
||| | : : : : : ||| | : : : : :
QY 137 GSSAISNGKNPHOREFENLSAEAVTRGVGMSTHTWCTPRIHP-PMESLPGIG--RPKL 193
||| | : : : : : ||| | : : : : :
DB 240 GSAAFMPQYPAQ-----GYGFGGAYTHMDLRRPQSPAPQOATPTTSHHSQP 288
||| | : : : : : ||| | : : : : :
QY 194 SNDAEDDKENLYSEAEER----LIGTSKEFEDESIRHTLVLSLQDA----YKDRQRI 245
||| | : : : : : ||| | : : : : :
DB 289 SNHSTSPAPANGVAFPARQVFTVGVSSTHTGNGSHSVPRRNDLIDLNHEDYSRV 348
||| | : : : : : ||| | : : : : :
QY 246 -----FRPLPLACHRLKNAPEYVHWSAENLHFSIYNDKQKKLFTLLTNHCRTRLALTG 300
||| | : : : : : ||| | : : : : :
DB 349 SVLEAFDPLNDNTGNTASDTSIYAEYDFDFLYSGD-----AATQYSDPMYEAVN 401
||| | : : : : : ||| | : : : : :
QY 301 GYEKKIGAAEVR-NLLATR-----NPSQLDSYIMA-----KVYVLASGAIGNP----- 343
||| | : : : : : ||| | : : : : :
DB 402 RWDKTATVSPNVGLIGWRQDFLSQSTSSSYQGVAPPEESLKLAEANGSGTISPPPLPP 461
||| | : : : : : ||| | : : : : :
QY 344 --QILYNSGFGSLQVLT-PRNDSLIPNLGRYITEOPMAFCOIVLRQ-----EFVDSVR 392
||| | : : : : : ||| | : : : : :

Db 462 RNQCYESQAAMPYVRPQQSVLTD--SYTSSIP---ANVYLDRRKTCRTLRYELISDQR 516

Qy 393 -DDPYGLPWKKEAVAHAKNPDTALP 418

Db 517 TDDPELLEFY-HWKEVRARYPHDDAP 542

RESULT 14

US-09-023-731-6
; Sequence 6, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawanura, Yukio; Morita,
; APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154

COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023.731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997

ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-023-731-6

Query Match 3.1%; Score 102.5; DB 4; Length 61;
Best Local Similarity 43.5%; Pred. No. 0.0027;
Matches 27; Conservative 6; Mismatches 24; Indels 5; Gaps 2;

Qy 104 FVNVKGLQOVSVVRNONVPTLDPGAWSPGSSA-----ISNGKNPHQREFNLSAE 159

Db 1 FVNIINGALQPISSPSTYQPTLAVAAW-APPIDPAEGQLVINGHNPQEAAGLNPGSA 59

Qy 160 VT 161

Db 60 VT 61

RESULT 15

US-08-549-515-5
; Sequence 5, Application US/08549515
; Patent No. 6054123
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Haemophilus Influenzae

; TITLE OF INVENTION: Dimethylsulphoxide Reductase Enzyme
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 Unviersity Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/549,515
; FILING DATE: 27-OCT-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-522
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1153
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 806 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-549-515-5

Query Match 3.1%; Score 100.5; DB 3; Length 806;
Best Local Similarity 18.6%; Pred. No. 0.42;
Matches 69; Conservative 56; Mismatches 139; Indels 107; Gaps 17;

Qy 31 YGNDGVDFIAGSGPIGATYAKLCVE----- 56

Db 230 FGNPFAETRMGG---GLTY---CIEQAKARSNAKMIIDPRYNDTGAGREDEWIPRPG 283

Qy 57 -----AGLRVVMVEIGAAD----SFYAVNAEETGTAVPYVP--GYHKKEIEFQKIDIRF 104

Db 284 TDAALVAALAVYMIQENLVDPQFLDKYCVGYDEKTLPADAPKNGHYKAVILGYGN--DGI 341

Qy 105 VNVKGLQOVSVVRNONVPTLDPGAWSPGSSAISNGKNPHQREFNLSAEAVTRGV 164

Db 342 AKTPEWAAKITGIPA--ERIIKLAREIGSTKP--AFISQGWGPQRRS-----NGELISR 393

Qy 165 -----GGMSTHTWCSTPRTHPPMESLPGIGRPPKLSNDPAEDDKENWELYSEAEERLIGT 217

Db 394 AMLPILTNVGIHGGNTGAR--ESAYSIPFVPMPTLK-----NPMKASIPMFLGT 441

Qy 218 STKEFDESIRHTLVLRSLQADYKQRIERFLPLACHRLKNAPEYVEVHSAENLFHSYN 277

Db 442 -----DAIIRGT-EMTALTDGIRGVDKLSPPIKVIVNYSNC--LINHOAQINRTHDILQ 493

Qy 278 DDQKKKFTLLTNHRCTRLALTGGYEKKIGAAEVNLLATRNPSQ-----LDSTYIMAKV 332

Db 494 DDTCCEMITIDNHMTS-----TAKYSDILLPDCPTSEOMDFALDAFVSNMA 540

Qy 333 YVLASGAIGNP 343

Db 541 YVIFADQVIKP 551

Search completed: September 3, 2002, 15:57:29
Job time: 3139 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 14:57:41 ; Search time 82.18 Seconds
(without alignments)
835.284 Million cell updates/sec

Title: US-09-856-327-2

Perfect score: 3284

Sequence: 1 MSLSFEQMLRDPMSQING.....IINTLKGGTGKNTGEHRNL 618

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3284	100.0	618	22	Lyophyllum shimeji
2	1232.5	37.5	623	20	Coriolus versicolor
3	1232	37.5	633	21	Pleurotus cornucop
4	1218.5	37.1	623	17	Pyranose oxidase.
5	1210	36.8	622	22	Trametes hirsuta p
6	1091	33.2	566	20	An antitumour prot
7	1044	31.8	564	21	T. matsutake pyran
8	1044	31.8	564	21	Trichoderma deri
9	200	6.1	544	20	G. oxydans D-sorbi
10	192	5.8	51	20	Peptide derived fr
11	191.5	5.8	615	20	Gluconate dehydro

12	177	5.4	67	21	AA10459	T. matsutake pyran
13	177	5.4	67	21	AA181951	N-terminal fragmen
14	168	5.1	34	22	AA197034	Lyophyllum shimeji
15	162	4.9	30	22	AA197031	Lyophyllum shimeji
16	159.5	4.9	657	21	AA15525	Arabidopsis thalia
17	159.5	4.9	675	21	AA15524	Arabidopsis thalia
18	159.5	4.9	748	21	AA15523	Arabidopsis thalia
19	142	4.3	29	22	AA197033	Lyophyllum shimeji
20	130.5	4.0	58	20	AA187540	Peptide derived fr
21	121	3.7	24	22	AA197032	Lyophyllum shimeji
22	121	3.7	704	20	AA131754	Candida tropicalis
23	120.5	3.7	546	20	AA128581	Choline oxidase (C
24	119	3.6	500	22	AA1869084	A. tumefaciens 1,5
25	119	3.6	545	21	AA136283	Arabidopsis thalia
26	119	3.6	727	21	AA136282	Arabidopsis thalia
27	119	3.6	746	21	AA136281	Arabidopsis thalia
28	119	3.6	2639	22	AB115016	Novel human diagno
29	118.5	3.6	785	21	AA182220	Humicola insolens
30	118.5	3.6	785	22	AA182220	H. insolens DSM 18
31	116.5	3.5	698	20	AA131752	Candida cloacae fa
32	115	3.5	50	20	AA187543	Peptide derived fr
33	110	3.3	551	21	AA135598	Streptomyces globi
34	108	3.3	35	20	AA187544	Peptide derived fr
35	107.5	3.3	964	22	AA130811	Amino acid sequenc
36	106.5	3.2	778	20	AA135090	Chlamydia pneumoni
37	106	3.2	572	22	AA182940	S. epidermidis ope
38	106	3.2	1356	22	AB167291	Drosophila melanog
39	105	3.2	609	22	AA156302	Propionibacterium
40	105	3.2	833	22	AA1868116	Drosophila melanog
41	104.5	3.2	1876	18	AA138757	Phosphatidyl inosi
42	104.5	3.2	1876	22	AB161970	Drosophila melanog
43	104.5	3.2	1876	22	AB166876	Drosophila melanog
44	103	3.1	437	22	AA129360	Novel human diagno
45	103	3.1	1686	19	AA170991	Human class II P13

ALIGNMENTS

RESULT 1

AA197035
ID: AA197035 standard; Protein; 618 AA.
XX AA197035;
XX
XX 20-JUL-2001 (first entry)
XX
XX Lyophyllum shimeji antibacterial protein.
XX
XX Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;
XX Pyricularia oryzae; Rhizoctonia solani; rice pathogen.
XX Lyophyllum shimeji.
XX
XX WO200121657-A1.
XX
XX 29-MAR-2001.
XX
XX 20-SEP-2000; 2000WO-JP06404.
XX
XX 21-SEP-1999; 99JP-0267238.
XX
XX (NISB) JAPAN TOBACCO INC.
XX (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
XX
XX Takakura Y, Kuwata S, Inoue Y;
XX
XX WPI; 2001-281598/29.
XX N-PSDB; AAF99980.
XX
XX Antibacterial protein and encoded gene isolated from Lyophyllum
XX shimeji, with activity against plant pathogenic bacteria, applicable in
XX agriculture e.g. rice cultivation at low concentration, produced at low

PT cost on large scale -

PS Claim 3; Page 43-45; 52pp; Japanese.

CC The present sequence is an antibacterial protein from the fungus
CC *Lyophyllum shimeji*. The protein was obtained from a fraction prepared
CC by extracting *Lyophyllum shimeji* with water and subjecting the extract
CC to ammonium sulphate precipitation. The protein inhibits the growth
CC of the plant pathogenic bacteria *Pseudomonas* and *Rhizoctonia*
CC *solani* at a relatively low concentration. *P. oryzae* and *R. solani* are
CC causative of the two major diseases of rice. The protein contains
CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.
CC The antibacterial protein can be produced at low cost on a large scale.

XX Sequence 618 AA;

Query Match 100.08; Score 3284; DB 22; Length 618;

Best Local Similarity 100.08; Pred. No. 2.1e-299;
Matches 618; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSLSTEOMLRDPRGMOINGQIPKNAIHEITYGNDGVDFIAGSGPIGAYAKLCVEAGLR 60

Db 1 mslsteqmlrdprsmqngqpknaihetygndgvdfiagsgppgatyaklcveaglr 60

QY 61 VYMVEIGAADSYAVNAEGTAVPYVPGYHKKNEIEFFQKIDRFVNVIKGALQQVSPVVR 120

Db 61 vymveigaadsyavnaeegtavpyvpgyhhkneiefqkldrfvnvirkalgqvsvpr 120

QY 121 NONVTLDPGANSAPPGSSAISINGKNPHOREFENISAEAVTVGVGMSTHTWCSTPRIHP 180

Db 121 nqnvptldpgawsappgssaisangknphqrefenisaeavtrvgvgmsthtwcstprihp 180

QY 181 PMSLPGIGRKLSDNDPAEDKRWENLYSEAEKLTGSTKPEDESIRHTLVLSLQDAYK 240

Db 181 pmslpgigrklslndpaedkewenlyseaeerlgtstkefdesirhtlvlsrqdayk 240

QY 241 DRQIFRPLPLACHRLKNAPEYVHSAENLPHSIYNDQKQKLFLLTNHRCRLALTG 300

Db 241 drqifrlplachrlknapeyvhhsaenlfhslynddkqkfltltnhrcrlalgtg 300

QY 301 GYEKIGAEVNNLILATRNPSQLDSYIMAKYVVLASGAINQPILYNSGSLQVTPRN 360

Db 301 gyeckigaeavnllatrnpsqldsyimakvylvlasgaignpqilynsqfslqvtpn 360

QY 361 DSLIPNLGRYITEQPMAFQCIIVLROEFVDSVRDDPYGLPWWKEAVAQHIKAKNPTDALPIP 420

Db 361 dsllpnlgryteqpmafqciivlrgefvdsvrddpyglpwkewaqlaknptdalp 420

QY 421 FRDPEFQVTPPTTEHPWHQTQIHRDAFSYGAVGPEVDSRVIVDLRWFGATDPEANNLLVF 480

Db 421 frdpeqvtpptteehpwhqtqihrdafsygavgpvdsrvivdlrwfgatdpeannllvf 480

QY 481 QNDVQGYAMPQPTFRYRPRSTASNVARKMMADMCVEASNLGGYLPSPQPMDFGLALH 540

Db 481 qndvqgyampqptfryrprstasnvrarkmmadmcveasnlggylpseppqfmdpglalh 540

QY 541 LAGTTRIGFDKATTVADNNLSVDFANLYVAGNGTIRTFGFGENPTLTSCHAIKARSII 600

Db 541 lagttrigfdkattvadnnlsvdfanlyvagngtirtfgfgenptltsmchaikarsii 600

QY 601 NTLKGGTGKNTGEHNL 618

Db 601 ntlkgtgdkntgehnl 618

RESULT 2

AAW94308

ID AAW94308 standard; protein: 623 AA.

XX AAW94308;

XX AAW94308;

DT 08-APR-1999 (first entry)

XX

DE Coriolus versicolor pyranose oxidase.

XX

KW Coriolus versicolor; modified: pyranose oxidase; heat stability;
KW pH stability; glucose; glucosone.

XX

OS Coriolus versicolor.

XX

PN JP11009271-A.

XX

PD 19-JAN-1999.

XX

PF 24-JUN-1997; 97JP-0181865.

XX

PR 24-JUN-1997; 97JP-0181865.

XX

PA (KIKK) KIKKOMAN CORP.

XX

DR WPI; 1999-145891/13.

XX

PT New modified pyranose oxidase gene and protein - useful for
PT efficient recombinant production of pyranose oxidase with excellent
PT Km, heat and pH stability

XX

PS Claim 1; Page 7-9; 10pp; Japanese.

XX

CC The present sequence represents pyranose oxidase derived from *Coriolus*
CC versicolor, which can be modified by adding, deleting or replacing at
CC least one amino acid (preferably where the amino acid at position 542
CC is replaced) and still retains pyranose oxidase activity. The modified
CC pyranose oxidase with the amino acid at position 542 replaced, has the
CC following properties: (a) Action: It oxidises glucose to glucosone;
CC (b) Substrate specificity: It reacts specifically with glucose and also
CC reacts with galactose, L-sorbose, D-xylose and 1,5-anhydro-D-glucitol;
CC (c) Stable pH: 3.5 to 11.0 at 50 degrees Celsius for 30 minutes; (d)
CC Optimum pH: Near 6.5; (e) Optimum temperature: Near 55 degrees Celsius;
CC and (f) Temperature stability: Stable up to 55 degrees Celsius. The
CC present invention also describes a method for the preparation of a
CC modified pyranose oxidase. The method efficiently prepares a modified
CC pyranose oxidase with excellent Km value, heat stability and
CC pH stability.

XX Sequence 623 AA;

Query Match 37.5%; Score 1232.5; DB 20; Length 623;

Best Local Similarity 45.0%; Pred. No. 1.7e-106;

Matches 273; Conservative 73; Mismatches 176; Indels 85; Gaps 16;

QY 37 DYFIAGSGPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEGTAVPYVPGYHKKNEIE 96

Db 48 dvivvgpgigctyarelveagykvmfdigeidsglki-----gahkntve 95

QY 97 FKQIDRFVNVTKGALQQVSVPRNQNVPTLDPGAWSPGSSAISNGKNPHOREFENUS 156

Db 96 yqknldkfnnvlgqqlmsvvpvntlvtdltsptswqa--ssffvrngsnpedplnls 153

QY 157 ABAVTRGVGGMSTHTWCSTPTTHPPMESLPGIGRKLSDNDPAE-DDKRWENLYSEAERLI 215

Db 154 gqavtrvgvgmsthtwcstprtdreq-----rpllvkddqdaadaewdrlytkaesyf 206

QY 216 GTSTKEFDESIRHTLVLSLQDAYKDRQIFRPLPLACHRLKNAPEYVHSAENL--H 273

Db 207 ktgtqgkesirhnvlvnlklaeeykg-qrdfqgiplaatr--rsptfveasantvfdlq 263

QY 274 SIYNDQKQKLFLLTNHRCRLALTGGEYKKGKIGAEVNNLILATRNPSQLDSY----- 327

Db 264 nrpntdapnerfnlfpavacerv-----vrn-----tsnseieslhlhdl 304

QY 328 -----IMAKYVVLASGAINQPILYNSGSLQVTPRN-DSLIPNLGRYITEQPMAPC 379

Db 305 egdrfeikadvilvtagavhnaqlvlnsgfqlgrpdpnanpqlpipsyigteisvfc 364

```

QY 380 QIVLRQEFVDSVRDDPY--GLP-----WKEAVAQHIKAKNPTDAL 417
Db 365 qtmstmlidsvksdmliirgnpgdglvsyvtgpaetknkhpwwnekvknhmmqhgedpl 424
QY 418 PIPRPDPEQVTPPTEEHQWHTQIHRDAFSYGAVGPEVDSRVIVDLRWFSGATDPEANNL 477
Db 425 pipfedpeqvvtlfgshpwhqihrdafsygavqsidsrlivdwrfgrtepkkeen 484
QY 478 LVFQNDVQDGYSPQPTFRYR--PSTASNVARKMMADCEVASNLGGYLTSPQFMDPG 536
Db 485 lwfsdkitdtyrnpqptfdrfagrtkskeadmmtdmcmvsakigglpgslpqfmeqg 544
QY 537 LAHLACTTRIGDOKA--TVADNNSLVWDFANLYVAGNCTIRTGPOENTLFSMCHAIK 594
Db 545 lvhlqgthrmgfdgedekccvntdsrvfgknlfgcgcnptayganptlamslaik 604
QY 595 SARSIIN 601
Db 605 sceyikn 611

RESULT 3
AAV52700
ID AAY52700 standard; Protein; 633 AA.
XX
AC AAY52700;
XX
DT 07-MAR-2000 (first entry)
XX
DE Pleurotus cornuopiae antitumour protein.
XX
KW Antitumour; cancer; tumour; treatment; expression; tumour suppressor;
KW p53; pBR; ss.
XX
OS Pleurotus cornuopiae.
XX
PN JP11315096-A.
XX
PD 16-NOV-1999.
XX
PF 07-AUG-1998; 98JP-0236349.
XX
PR 08-AUG-1997; 97JP-0215311.
PR 02-MAR-1998; 98JP-0066176.
XX
PA (NEW-) NEW FOOD CREATION GIJUTSU KENKYU KUMIAI.
XX
DR WPI; 2000-058170/05.
DR N-PSDB; AAZ46411.
XX
PT An antitumour protein derived from Pleurotus cornuopiae and its gene -
PT useful for treatment of cancer including those caused by abnormal
PT expression of cancer inhibitory gene (e.g. p53 and pBR)
XX
PS Claim 3; Page 13-15; 23pp; Japanese.
XX
CC The invention relates to a novel antitumour protein extracted from
CC fruiting bodies of the fungus Pleurotus cornuopiae. The protein and
CC nucleotides encoding it are useful for the treatment of cancer,
CC including those caused by abnormal expression of tumour suppressor
CC genes such as p53 and pBR. This sequence represents the antitumour
CC protein.
XX
SQ Sequence 633 AA;

Query Match 37.5%; Score 1232; DB 21; Length 633;
Best Local Similarity 44.6%; Prod. No. 2e-106;
Matches 275; Conservative 77; Mismatches 193; Indels 72; Gaps 18;

QY 18 INQIIPKNAHEHYGNDGVDFYTAGSGPIGATYAKLCVEAGLRVWVVEIGAADSFYAVNA 77
Db 44 ipdklpsg-----dtskfdvrvigspvgstyarlliveagfkvmfeligdeids----- 91

```

```

QY 78 EESTAVYVPVGYHKKNEIEFQKIDREVNVIKQALQOVSVFVRNQNVPTLDPGAWSPGP 137
Db 92 --grkl-----gshkntveyknlidkfvhviqglmpsvpnkyvadtispaswqa--s 143
QY 138 SSALSNCKNPHQREFFENLSAEAVTRGVGGSTHTWCTSTPRIHPPMESLPGIGRPKL-SND 196
Db 144 thvrnganpeqdpftnlsggavtrvggmthwctcatprfhks-----erpklvkdd 196
QY 197 PAEDDKENNELYSAERLIGTSTKEFOESIRHTLVLSLQDAYKDRIRFRLPLACHRL 256
Db 197 dsadatwerlydiaesfvktghnqfogsirhmlvleklgesyag-qrgfeqiplaaqrt 255
QY 257 KNAPEYVEHWSAENLF--HSIYNDKQKQLFTLLTNHRCRTLALTGGYKKGIAAEVRNL 314
Db 256 n--prfvevssahvtfdlenrpnaddegrfnlfpavvcervt-rdsldrkievhdh 312
QY 315 LATRNPSSQLDSY--IMAKVYVVLASGAIGNPQILYNSGSLQVTPRND-----SLIPNLGR 369
Db 313 lsg-----drykvkadvfilcsgavhmpqilvnsqf-grmgqpdsslpptlilpylgs 364
QY 370 YITEQPMACQIVLRQEFVDSVRDDP--YGLP-----WKEAVAQ 407
Db 365 yiteqtltfcqvfstelnlvksdmliivtppgdydvstftpdspnknkhpwnwnekvqk 424
QY 408 HIAKNPTDALPIPRDPEQVTPPTEEHQWHTQIHRDAFSYGAVGPEVDSRVIVDLRWF 467
Db 425 hnmqhqedplpiddpeqvvtlfgdthpwhqihrdafsygavaesidsrlvvdwrff 484
QY 468 GATDPEANNLVFPONDVQDGYSPQPTFRYR--PSTASNVARKMMADCEVASNLGGYLP 526
Db 485 grtepvveenklwfskqitdaynlpqtfstfrpqrtaqeaelmmadmctmstkvggflp 544
QY 527 TSPQFMDPGLALHLAGTTRIGRDKATTVA--DNNSLVWDFANLYVAGNCTIRTGFCGNP 584
Db 545 gsyppqfmapglvhlhgghrmgfddeadkacvdcnkvfmgmenifggcgngitayasn 604
QY 585 TLTSMSCHAIKSARSIIN 601
Db 605 tltavalairskyirn 621

RESULT 4
AAV99628
ID AAR99628 standard; Protein; 623 AA.
XX
AC AAR99628;
XX
DT 27-NOV-1996 (first entry)
XX
DE Pyranose oxidase.
XX
KW Pyranose oxidase; glucose; oxidation; glucanase; assay; diabetes;
KW marker; diagnosis; L,5-anhydro-D-sorbitol.
XX
OS Coriolus versicolor.
XX
PN DE19545780-Al.
XX
PD 13-JUN-1996.
XX
PF 07-DEC-1995; 95DE-1045780.
XX
PR 24-MAY-1995; 95JP-0124835.
PR 07-DEC-1994; 94JP-0304086.
XX
(KIKK ) KIKKOMAN CORP.
XX
PA Kawai G, Koyama Y, Minamihara T, Nishimura I, Okada K;
PI Suzuki M;
XX
DR WPI; 1996-278990/29.
DR N-PSDB; AAT34420.

```

XX DNA encoding protein with pyranose oxidase activity at neutral pH -
PT useful for the determination of glucose in body fluids or foods, or
PT 1,5-anhydro-D-sorbitol used as marker for diabetes diagnosis
XX
PS Claim 1; Page 14-16; 22pp; German.
XX
CC The present sequence encodes a protein isolated from *Corioliolus versicolor*,
CC which has the enzyme activity of pyranose oxidase (PO). The PO oxidises
CC glucose to gluconone and has an optimum pH of 7-7.5. It has a mol. wt. of
CC 290000 (determined by gel filtration) and is stable at around 50deg.C.
CC PO can be used for measurement of glucose in, e.g. foods or body fluids,
CC or 1,5-anhydro-D-sorbitol which is an important marker used in the
CC diagnosis of diabetes.
XX
SQ Sequence 623 AA;

Query Match 37.1%; Score 1218.5; DB 17; Length 623;
Best Local Similarity 44.8%; Pred. No. 3.5e-105;
Matches 272; Conservative 72; Mismatches 178; Indels 85; Gaps 16;

QY 37 DVFIAGSPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEEGTAVPVPGYHKKNEIE 96
Db 48 dvlvvgspigctyarelveagkykvmfdgldsglki-----gahkknktve 95

QY 97 FOKDIDRVNVIKALQOVSPVRNQNVPTLDPGAWSPGSSAISNGKNPHOREFENLS 156
Db 96 yqknldkfvnvlgqqlmsvpyntlvdtlspstswqg--ssffvrngsnpeqdlrnl 153

QY 157 ABATRVGGSTHTWCTSPRIHPHPEMLSGIGRPKLSNDPAE--DKENNELYSAEERLI 215
Db 154 gqavtrvggmsthwctatprdfreq-----rpllvkddqaddaewdrlytkaesyf 206

QY 216 GTSTKEFDESIRHTLVLSLQDAYKDRIRFPLPLACHRLKNAPEYVHWSAENLF--H 273
Db 207 krtgdtqfkesirhnlvlkklaeeykg-grdfqqlplaatr--rsptfvevssatvdlq 263

QY 274 STYNDKOKKLTLLTNHRCRTRLATLGGYEKKIGAAEVRNLLATRNPSQLDSY----- 327
Db 264 nrpndapnerfnlfpavacerv-----vtn-----tsnseieslhihdl 304

QY 328 -----INAKYVYLASGAIQNPQILYNSGFSGL-QVTPRN--DSLIPNLGRVITEQPMFAC 379
Db 305 sgdrfeikadvfvtagavhnaqlvnsfgglgrpdpbanppqlipsrlytedgslvfc 364

QY 380 QIVLRQEFVDSVRDDPY--GLP-----WKEAVAQHTAKNPTDAL 417
Db 365 qtvmtelidskdmilrgnpgdlgyvtytpgaetkhkdpdwnekvknhmmqhqedpl 424

QY 418 PIPFRDPEQVPTTPEEHPHHTQIHRDAFSYGAVGPEVDSRVIVDLRWFATDPEANNL 477
Db 425 pipfedpeqvvtlfgqshpwhqtqhrdafsygavqqtldsrllvdrffgtrtepeenk 484

QY 478 LVFQNDVODGYSMPQTPRYR--PSTASNVRARKMADCEVASNLGGYLPSPPOFMDPG 536
Db 485 lwfcdkitdtynmqptdfdfpagrtskaedmmtdmcmvkwakigglfpgslpqlfmgpg 544

QY 537 LALHLAGTRIGFEDKA--TTVADNNLSVDFANLYVACNGTRTGFGNGENPILTSCHAIK 594
Db 545 lylhlggthrmgfdedeqdeccvntdsrvfgknlfigcgcnlptayganpiltamslaik 604

QY 595 SANSIIN 601
Db 605 sceyikn 611

RESULT 5
ID AAB48832
XX AAB48832 standard; Protein; 622 AA.
AC
XX AAB48832;

DT 13-MAR-2001 (first entry)
XX
DE Trametes hirsuta pyranose oxidase.
XX
KW Pyranose oxidase; expression construct; recombinant production;
KW monosaccharide oxidation; 2-keto derivative;
KW hydrogen peroxide production.
OS Trametes hirsuta.
PN US6146865-A.
XX
PD 14-NOV-2000.
XX
PF 05-MAY-1999; 99US-0305381.
XX
PR 08-JUN-1998; 98DK-0000774.
PR 10-JUN-1998; 98US-0088724.
XX
PA (NOVO) NOVO NORDISK AS.
XX
PI Schneider P, Christensen S, Lassen SF;
XX
DR WPI; 2001-049055/06.
DR N-PSDB; AAC87518, AAC87519.
XX
XX Novel nucleic acid molecule encoding polypeptide having pyranose
PT oxidase activity used to design oligonucleotide probes to identify and
PT clone DNA encoding the polypeptide from different genera or species -
XX
PS Claim 5; Fig 1; 20pp; English.
XX
CC The invention relates to nucleic acids (e.g., AAC87518, AAC87519) which
CC encode Trametes hirsuta pyranose oxidase (AAB48832). The invention also
CC relates to expression constructs, expression vectors and recombinant
CC cells comprising pyranose oxidase nucleic acid sequences, and the
CC recombinant production of Trametes hirsuta pyranose oxidase. Pyranose
CC oxidase catalyses the oxidation of several monosaccharides in the
CC pyranose form at position C2 to produce 2-keto derivatives with the
CC release of hydrogen peroxide. Nucleic acids encoding Trametes hirsuta
CC pyranose oxidase may be used to produce the enzyme and to design
CC oligonucleotide probes to identify and clone genomic pyranose oxidase
CC cDNA or genomic DNA from different genera or species of microorganisms
CC (fungi or bacteria). The present sequence represents pyranose oxidase
CC from the fungus Trametes hirsuta.
XX
SQ Sequence 622 AA;

Query Match 36.8%; Score 1210; DB 22; Length 622;
Best Local Similarity 44.6%; Pred. No. 2.2e-104;
Matches 274; Conservative 71; Mismatches 197; Indels 72; Gaps 16;

QY 37 DVFIAGSPIGATYAKLCVEAGLRVVMVEIGAADSFYAVNAEEGTAVPVPGYHKKNEIE 96
Db 48 dvlvvgspigctyarelveagfnvamefdgldsglki-----gshkknktve 95

QY 97 FOKDIDRVNVIKALQOVSPVRNQNVPTLDPGAWSPGSSAISNGKNPHOREFENLS 156
Db 96 yqknldkfvnvlgqqlmsvpyntlvdtlspstswqg--stffvrngsnpeqdlrnl 153

QY 157 ABATRVGGSTHTWCTSPRIHPHPEMLSGIGRPKLSNDPAE--DKENNELYSAEERLI 215
Db 154 gqavtrvggmsthwctatprfek-----lqrpvlvknkskaddaewdrlytkaesyf 206

QY 216 GTSTKEFDESIRHTLVLSLQDAYKDRIRFPLPLACHRLKNAPEYVHWSAENLF--H 273
Db 207 krtgdtqfkesirhnlvlkklaeeykg-vrdfqqlplaatr--qsptfvevssatvdl 263

QY 274 STYNDKOKKLTLLTNHRCRTRLATLGGYEKKIGAAEVRNLLATRNPSQLDSYIMAKVY 333
Db 264 nrpndapkrfnlfpavacnvrndnanselv9-lvdrldnggkslt-----lkakvy 316

CC Pyricularia oryzae or Thanatephorus cucumeris and has a M.W. of ca.
 CC 50 kD by gel filtration and shows the presence of components of ca.
 CC 15 kD and ca. 15 kD by SDS-PAGE and maintains the above antibacterial
 CC activity by being heated at 60 degrees C for 10 minutes in a neutral
 CC aqueous solution and in which the above antibacterial activity is
 CC inactivated by being heated at 80 degrees C for 10 minutes in a neutral
 CC aqueous solution. The invention also describes a method for the
 CC determination of pyranose in a sample in which the above reagent for the
 CC determination of pyranose is reacted with an enzyme in the sample and the
 CC hydrogen peroxide formed is reacted with an enzyme to develop a color, a
 CC method for diagnosing a disease accompanied by abnormality in sugar
 CC metabolism in which the above reagent for the determination of pyranose
 CC is reacted with pyranose in the sample and the hydrogen peroxide formed
 CC is reacted with an enzyme to develop a color, and a kit for pyranose
 CC analysis or the diagnosis of a disease accompanied by abnormality in
 CC sugar metabolism. The pyranose oxidase can be used as a diagnostic agent
 CC for diabetes mellitus. This sequence represents the Tricholoma matsutake.
 CC pyranose oxidase protein which is described in the method of the
 CC invention.
 XX
 SQ Sequence 564 AA;

Query Match 31.8%; Score 1044; DB 21; Length 564;
 Best Local Similarity 40.1%; Pred. No. 7.7e-89;
 Matches 243; Conservative 94; Mismatches 197; Indels 72; Gaps 17;

QY 17 QINGOIPKNAIHEHYGNDGV-----DVFIAGSGPIGATYAKLCVE--AGLRVVMVEIGAAD 70
 DB 9 kindllqrsqgdltsqdeivhytdvfiagsgpiactyarhiidntsttkvymaelsqsd 68

QY 71 SFYAVNAEEGTAVPVPGYHKHNEIEFQKIDRFVNVIKGALQOVSPVRNQNVTLPDGP 130
 DB 69 n-----pvi-gahhrnsikfkqtdkfvniingalqpsispsdyqptlava 115

QY 131 AWSAPPGSSA-----ISNGKNPHQREFENLSAEAVTRGVGGMSTHWTCTPRTHPPMESLIP 186
 DB 116 aw-appldpaegqlvimghnpqaeaglnlpgsavtrtvggmawhwtcacptphd----- 168

QY 187 GIGRKLNDPAEDDKENWELSEARLIGTSTKEFDSIRHTLVLSLQDAYKDRQRF 246
 DB 169 -----eervnnpv-dkqefdaalleraktlnvhsdyddsiqvivketlqtl-dasrgv 222

QY 247 RPLPLACHRLKNAPYEVHWSAENLFHSIYNDKQKLFLLTNHRCRLALTGGERKI 306
 DB 223 tclplgvertdnplyvtgad----tvlgdvkpsprfvlvtetrckfivsetnptqv 278

QY 307 GAAEVRNLLATNPSSQLDSYIMAKVYVLASGAIGNPQILYNSGFGSLQVTPRNDSLIPN 366
 DB 279 vaallrnl-----ntsn--delvvaqsviaagavctpqilwns-----nirph-----a 321

QY 367 LGRYITEQPMAPFCQIVLRQEFVSDVRDDPYGLPMWKEAUAOHIKAKNPDTDALPIPRDPEP 426
 DB 322 lgrlysegsmtfcqlvrlsrldvsiatdpr-----faakveahkknhdvlpipthepep 377

QY 427 QVTPTTEHPHMHQIHRDAFSYAGVGEVDSRVIVDLRWFGCATDPEANNLVFE----- 480
 DB 378 qmlyptsdtsfwhvqvrh--yafgvgpkadprvvvdrlrfgkgsdiveenrvtfgnpkl 435

QY 481 ---QNDVQDGYSMPOPTFRYRSTASNVARKMADMCVEASNLGGYLPSTPPQPMDEL 537
 DB 436 rdweagvtdtygmpqptfhvkrtnadgdrdgrmndmtnvanilggylpgsypqfmapgl 495

QY 538 ALHLAGTTRIGFDKATTVADNNLSVWDFANLYVAGNGTIRTGFGFNPFLTSMCHAIKASR 597
 DB 496 aqhikgttrigtddqtsvadptskvnmfdnlwvgngcldpatacntrtsvayaikgae 555

QY 598 SIINTL 603
 DB 556 avvsyl 561

RESULT 8

AAV81952
 ID AAV81952 standard; Protein; 564 AA.
 XX
 AC AAV81952;
 XX
 DT 07-JUL-2000 (first entry)
 XX
 DE Trichoderma derived antifungal protein sequence.
 XX
 KW Antifungal protein; Pyricularia oryzae; Rhizoctonia solani; mushroom;
 KW growth inhibitor; plant pathogenic fungi; antibacterial agent;
 KW N-terminal fragment.
 XX
 OS Trichoderma matsutake.
 PN WO200014242-A1.
 XX
 PD 16-MAR-2000.
 XX
 PF 19-AUG-1999; 99WO-JP04441.
 XX
 PR 08-SEP-1998; 98JP-0270606.
 XX
 PA (NITSB) JAPAN TOBACCO INC.
 PA (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
 XX
 PI Takakura Y, Kuwata S, Ohta S;
 XX
 DR WPI; 2000-256990/22.
 DR N-PSDB; AAA07403.
 XX
 PT Mushroom-derived antibacterial protein against plant pathogenic fungi
 PT of rice, with activity and thermal stability, obtainable cheaply on
 PT large scale, useful in agriculture
 XX
 PS Claim 3; Page 41-45; 52pp; Japanese.
 XX
 CC This sequence is the Trichoderma matsutake antibacterial protein of
 CC the invention. The protein has activity against at least Pyricularia
 CC oryzae and Rhizoctonia solani, and is obtained from a fraction of an
 CC aqueous extract of a mushroom precipitated by the ammonium sulphate
 CC precipitation method. The protein has a molecular weight of about 210 kD
 CC as determined by the gel filtration method, includes components of about
 CC 15 kD and 50 kD in SDS-PAGE, and is stable to heating in an aqueous
 CC neutral solution at 60 degrees C for 10 minutes but with loss of
 CC antibacterial activity after heating in the solution at 80 degrees C for
 CC 10 minutes. The protein is used for inhibiting the growth of plant
 CC pathogenic fungi e.g. Pyricularia oryzae and Rhizoctonia solani.
 CC It is useful in treating rice plants, and is applicable in agriculture as
 CC an antibacterial agent. The protein has activity at relatively low
 CC concentrations, and can be produced at low cost on large scale.
 XX
 SQ Sequence 564 AA;

Query Match 31.8%; Score 1044; DB 21; Length 564;
 Best Local Similarity 40.1%; Pred. No. 7.7e-89;
 Matches 243; Conservative 94; Mismatches 197; Indels 72; Gaps 17;

QY 17 QINGOIPKNAIHEHYGNDGV-----DVFIAGSGPIGATYAKLCVE--AGLRVVMVEIGAAD 70
 DB 9 kindllqrsqgdltsqdeivhytdvfiagsgpiactyarhiidntsttkvymaelsqsd 68

QY 71 SFYAVNAEEGTAVPVPGYHKHNEIEFQKIDRFVNVIKGALQOVSPVRNQNVTLPDGP 130
 DB 69 n-----pvi-gahhrnsikfkqtdkfvniingalqpsispsdyqptlava 115

QY 131 AWSAPPGSSA-----ISNGKNPHQREFENLSAEAVTRGVGGMSTHWTCTPRTHPPMESLIP 186
 DB 116 aw-appldpaegqlvimghnpqaeaglnlpgsavtrtvggmawhwtcacptphd----- 168

QY 187 GIGRKLNDPAEDDKENWELSEARLIGTSTKEFDSIRHTLVLSLQDAYKDRQRF 246
 DB 169 -----eervnnpv-dkqefdaalleraktlnvhsdyddsiqvivketlqtl-dasrgv 222

QY 247 RPLPLACHRLKNAPYEVHWSAENLFHSIYNDKQKLFLLTNHRCRLALTGGERKI 306
 DB 223 tclplgvertdnplyvtgad----tvlgdvkpsprfvlvtetrckfivsetnptqv 278

QY 307 GAAEVRNLLATNPSSQLDSYIMAKVYVLASGAIGNPQILYNSGFGSLQVTPRNDSLIPN 366
 DB 279 vaallrnl-----ntsn--delvvaqsviaagavctpqilwns-----nirph-----a 321

QY 367 LGRYITEQPMAPFCQIVLRQEFVSDVRDDPYGLPMWKEAUAOHIKAKNPDTDALPIPRDPEP 426
 DB 322 lgrlysegsmtfcqlvrlsrldvsiatdpr-----faakveahkknhdvlpipthepep 377

QY 427 QVTPTTEHPHMHQIHRDAFSYAGVGEVDSRVIVDLRWFGCATDPEANNLVFE----- 480
 DB 378 qmlyptsdtsfwhvqvrh--yafgvgpkadprvvvdrlrfgkgsdiveenrvtfgnpkl 435

QY 481 ---QNDVQDGYSMPOPTFRYRSTASNVARKMADMCVEASNLGGYLPSTPPQPMDEL 537
 DB 436 rdweagvtdtygmpqptfhvkrtnadgdrdgrmndmtnvanilggylpgsypqfmapgl 495

QY 538 ALHLAGTTRIGFDKATTVADNNLSVWDFANLYVAGNGTIRTGFGFNPFLTSMCHAIKASR 597
 DB 496 aqhikgttrigtddqtsvadptskvnmfdnlwvgngcldpatacntrtsvayaikgae 555

QY 598 SIINTL 603
 DB 556 avvsyl 561


```

XX OS Lyophyllum shimeji.
XX PN WO200121657-A1.
XX XX
XX PD 29-MAR-2001.
XX PF 20-SEP-2000; 2000WO-JP06404.
XX PR 21-SEP-1999; 99JP-0267238.
XX PA (NISB ) JAPAN TOBACCO INC.
XX PA (NORO ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
XX PI Takakura Y, Kuwata S, Inoue Y;
XX DR WPI; 2001-281598/29.
XX XX
XX PT Antibacterial protein and encoded gene isolated from Lyophyllum
PT shimeji, with activity against plant pathogenic bacteria, applicable in
PT agriculture e.g. rice cultivation at low concentration, produced at low
PT cost on large scale -
XX PS Claim 2; Page 32; 52pp; Japanese.
XX CC The present sequence is part of an antibacterial protein from the fungus
CC Lyophyllum shimeji. The protein was obtained from a fraction prepared
CC by extracting Lyophyllum shimeji with water and subjecting the extract
CC to ammonium sulphate precipitation. The protein inhibits the growth
CC of the plant pathogenic bacteria Pyricularia orizae and Rhizoctonia
CC solani at a relatively low concentration. P. orizae and R. solani are
CC causative of the two major diseases of rice. The protein contains
CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.
CC The antibacterial protein can be produced at low cost on a large scale.
XX SQ Sequence 30 AA;

Query Match          4.9%; Score 162; DB 22; Length 30;
Best Local Similarity 100.0%; Pred. No. 4.4e-08;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 NAEETAVPYVPGYHKNEIEFQKDIRFV 105
Db 1 naeegtavpyvpyghkneiefqkdidrfv 30

Search completed: September 3, 2002, 15:56:34
Job time: 3533 sec

XX OS Lyophyllum shimeji.
XX PN WO200121657-A1.
XX XX
XX PD 29-MAR-2001.
XX PF 20-SEP-2000; 2000WO-JP06404.
XX PR 21-SEP-1999; 99JP-0267238.
XX PA (NISB ) JAPAN TOBACCO INC.
XX PA (NORO ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
XX PI Takakura Y, Kuwata S, Inoue Y;
XX DR WPI; 2001-281598/29.
XX XX
XX PT Antibacterial protein and encoded gene isolated from Lyophyllum
PT shimeji, with activity against plant pathogenic bacteria, applicable in
PT agriculture e.g. rice cultivation at low concentration, produced at low
PT cost on large scale -
XX PS Example 2; Page 24; 52pp; Japanese.
XX CC The present sequence is part of an antibacterial protein from the fungus
CC Lyophyllum shimeji. The protein was obtained from a fraction prepared
CC by extracting Lyophyllum shimeji with water and subjecting the extract
CC to ammonium sulphate precipitation. The protein inhibits the growth
CC of the plant pathogenic bacteria Pyricularia orizae and Rhizoctonia
CC solani at a relatively low concentration. P. orizae and R. solani are
CC causative of the two major diseases of rice. The protein contains
CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.
CC The antibacterial protein can be produced at low cost on a large scale.
XX SQ Sequence 34 AA;

Query Match          5.1%; Score 168; DB 22; Length 34;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 AERLIGTSTKEFDESIRHTLVLRSLQDAYKDRQR 244
Db 1 aerlgtstkefdesirhtlvrlslrqdaykdrqr 34

RESULT 15
AAB97031
ID AAB97031 standard; Peptide; 30 AA.
XX AC AAB97031;
XX DT 20-JUL-2001 (first entry)
XX DE Lyophyllum shimeji antibacterial protein fragment #1.
XX KW Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;
XX KW Pyricularia orizae; Rhizoctonia solani; rice pathogen.
XX OS Lyophyllum shimeji.
XX PN WO200121657-A1.
XX PD 29-MAR-2001.
XX PF 20-SEP-2000; 2000WO-JP06404.
XX PR 21-SEP-1999; 99JP-0267238.
XX PA (NISB ) JAPAN TOBACCO INC.
XX PA (NORO ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
XX XX

```

Result No.	Query #			DB	ID	Description
	Score	Match	Length			
1	1142.5	39.4	623	3	P79076	P79076 coriolus ve
2	170.5	5.9	573	16	Q9P190	Q9P190 campylobact
3	170	5.9	551	2	Q9RH54	Q9RH54 pantoaea agg
4	166.5	5.7	579	16	Q9ATV6	Q9ATV6 caulobacter
5	159	5.5	615	2	O34214	O34214 pectobacter
6	158.5	5.5	529	17	Q9H0R8	Q9H0R8 halobacter1
7	157	5.4	545	2	Q5LBR8	Q5LBR8 gluconobact
8	148	5.1	553	2	Q9XCRO	Q9XCRO pantoaea cit
9	144	5.0	722	16	Q9RZ26	Q9RZ26 deinococcus
10	141	4.9	523	16	Q9HC76	Q9HC76 rhizobium l
11	141	4.9	591	16	Q91IK8	Q91IK8 pseudomonas
12	128.5	4.4	748	10	Q9M0H4	Q9M0H4 arabidopsis
13	128.5	4.4	748	10	Q94BP3	Q94BP3 arabidopsis
14	127	4.4	499	16	Q98BP2	Q98BP2 rhizobium l
15	117.5	4.1	620	3	Q12623	Q12623 humicola gr
16	116	4.0	502	5	O18672	O18672 caenorhabdi


```
QY 257 VYVLASGAINPQILYNSGSLQVTPRNDSLIPN-----LGRYTEQPMFAFCQIVLRQEF 312
Db 290 AFALANGCIETPRLL-----MAANDANPNGIANDMVGRRNMDHSGFCSFLTKE-- 341
QY 313 VDSVRDDPYGLPWKKEAVAQIAKNPT-DALPIPFROPEQVTPPTEEPHWHFOI----- 367
Db 342 -----PW-----LGKPAQSSCWGYRGD-----FRDYSANKVILNNI 377
QY 368 -----HRDAFSYGAVGPEVDSRV-----IVDLRWFAGTDPPEANNLLVFQNDVQGYSM 415
Db 378 SRVVTATQAAKMGVLGKALDEETRYRAVHSVDSLSISLEPLDPDENRLTSLKTRKDPHGL 437
QY 416 POFTRFRPSTASVNRARKMADCEVASNLGGVLPSPFPQFMDPGLAL--HLAGTTRIG 473
Db 438 PCPDIIYDVGIVYRKGAESAQAQLEHI-----GOLFKAKEFTISQGLNANNHNGVIMG 492
QY 474 FDKATTVADNNSLVDFANTYVAGNGVIRTFGFGENTPLTSMCHAISARSISINTLKG 530
Db 493 KNAKEAVVDCNCRADFHNELWLPGGGAIPASVVSNSLTMAALGLKAAHDISLRMG 549

RESULT 4
ID Q9A7T6 PRELIMINARY; PRT; 579 AA.
AC Q9A7T6;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE OXIDOREDUCTASE, GMC FAMILY.
GN CC1634.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Ulfenback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT *Complete genome sequence of Caulobacter crescentus.*;
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005839; AAK23612.1;
DR TIGR; CC1634;
DR Interpro; IPR000205; NAD_binding.
KW Complete proteome.
SQ SEQUENCE 579 AA; 64397 MW; 7125C4DFAD618F10 CRC64;
```

```
QY 89 VGGMSTHTWCTSTPRIHPMESLPGIGRPLKND--PAEDDKENNELYSEARLIGTSTK 145
Db 120 VGGSLTWGRQSYR-HSPID-FEANAREGIAVDPIRYEDLAPW---YEHVERFVGSQ 174
QY 146 EFDESIRH-----TLVLSRQDAYKDRQRI-FRPLPLACHRLKNAPEVVEHSA--- 193
Db 175 A--BGLPHFPDGHVQPPMELNCVEKAFKARSEARFPERRVTIGTAHLTDPTBQLAGR 232
QY 194 -----ENL-----FHSIYNDKQKFLTLTNHRCFRLALTGG-----YEKKIGAAE 235
Db 233 TKCOYRNLICRGCPFGAYYSNSG-----GLIAERTGNLVRPNISVTELIYDERAGRAS 288
QY 236 VRNLLATRNPSQLSDSIYAKVYVLASGAINPQILYNSGSLQVTPRNDSLIPN---- 291
Db 289 GVRIL----DAETRKDEEFHADVIFLCASALNSAWIMNS-----TSSRFPNGFN 335
```

```
QY 292 -----LGRYTEQPMFAFCQIVLRQEFVDSV-----RDDPYGLPWKKEAVAQIAKNPTDALP 343
Db 336 ASDQLGRNVMDHHLGAGATGOAPEADMYFSGRPNGIYVPRFRN-LGDAASKRSDYLRG 394
QY 344 IPRDPPEQVTPPTEEPHWHFOIHRDAFSYGAV-----GPEVDSRVIVDLRWFAT 395
Db 395 FGQGGAGRAT-----WERDRGGGRGFGAARKAALSQPGPW-----MGLSGFGEM 441
QY 396 DPEANNLLVFQNDVQGYSMPOPTFRYRPSSTASVNRARKMADCEVASNLGGVLPSTSP 455
Db 442 LPYADNRVTLNRDVEDFEGPLTLTNVTRMDNEMARMDMQAAAEMLAAGFONVRAHD 501
QY 456 QFMDPGLALHLAGTTRIGFGRATTVADNNSLVDFANLYVAGNGTIRTFGFGENTPLTSMC 515
Db 502 NGFAPGLGIHEMGTARMGRDPKTSVLNAHQVHECKNVYVTDGAAMASASCVNPSLTVMA 561
QY 516 HAKSARSISINTLKG 531
Db 562 LTAADAAHVAVRARRKG 577

RESULT 5
ID O34214 PRELIMINARY; PRT; 615 AA.
AC O34214;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE DEHYDROGENASE SUBUNIT PRECURSOR.
OS Pectobacterium cyripedii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=55209;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC29267;
RX Yum D.-Y., Lee Y.-P., Pan J.-G.;
RA "Cloning and expression of a gene cluster encoding three subunits of
RT membrane-bound gluconate dehydrogenase from Erwinia cyripedii ATCC
RT 29267 in Escherichia coli.";
RL J. Bacteriol. 179:6566-6572(1997).
DR EMBL; U97665; AAC45885.1;
KW SIGNAL.
FT CHAIN 1 22 DEHYDROGENASE SUBUNIT.
FT CHAIN 23 615
SQ SEQUENCE 615 AA; 67241 MW; B9E1A84FD035609A CRC64;
```

```
Query Match 5.5%; Score 159; DB 2; Length 615;
Best Local Similarity 19.9%; Pred. No. 8.3e-05;
Matches 124; Conservative 85; Mismatches 257; Indels 158; Gaps 32;
```

```
QY 4 EGTAIVYVPGYHKKNEIEFQKIDRFVNI-----KGALQOVSVPRNQNVPTLDPCAWS 58
Db 5 ERVSVP-VSGYSRGEGVTVADELKVDVAVVFGWAGAIMAKELTEAGLNVALERG--- 60
QY 59 APCSSAISNCKNP-----HQEFENLSAEVT-----R 87
Db 61 --PHRDYTPDGAYPQSIDELTYNIRKKLFQDLKSTVTIRHDSQTAVPYRQLAALPGLT 118
QY 88 GVGMSTHTWCTSTPRIHPMESLPG-----IGRPLKNDPAEDD--KEMNEL---YSEAE 137
Db 119 GTGAGLHWSGVHVRVDPVDELNLSHYEARVYKGFIEGCTIQDFGYSYNELEFPFOAE 178
QY 138 RLIGTSTKEFDESIRHTLVLSRQDAY--KDRQIRFRLP-----LACHRLKNAPEVVE 190
Db 179 KVFTCTSGSAW--TIKGMWIGKEKGNFVAPDRSSDF-PLPAKRTYSQAQLFAQAESVGY 235
QY 191 H-----SAENLPHSIYND-----KOKKLTFL 212
Db 236 HPYDMP SANTSGPYTNTYGAQMGPFCNFCGYSVACVYKASPNVNLPALROEPKFE 295
```

Qy	213	LTNHRCTRALTGGYEKKIGA-----AEVRNLLATRNPPSSQ---LDSYIMAKVYVVLASGAIG	266
Db	296	RNNAYVLRVNLGTGDKKRGATVYLDGGGREW---QPADLVILSAQFHNHVLMLLSGIG	352
Qy	267	NP-QILYNSGFSGLQVTPRNDLSLIPNL-GRYITEQPM---AFCQIVLQFEFVDSVRDDPY	321
Db	353	QPYNPITNEGVGVRNFAYQNISTLKALFDKNTTNPFIGAGGAGVAVDDFNADNFDHGPy	412
Qy	322	GL-----PMWKEAVAQHIAKNPTDALPIPFDPPEQVTPFTTEHPWHHTQIHRDAFSYGAV	377
Db	413	GFVGGSPFW---VNQAGTKPVSGLETPKGTN-----WGSQWK-----AAV	450
Qy	378	GPEVDSRVIVDLRFWGFATDEANLLVFQNDQDGYMSP--OPTERYRPTASNVYR-ARK	434
Db	451	ADTYNHHIISMDAH--GAHOSYRANYLDLPNKVYQGPLLRMTEDWDQ---NDIRMAQF	505
Qy	435	MMADMCEV--ASN---LGGYLPTSPQPMDFCL--ALHACTRTIGFDPKATTVADNNSL	486
Db	506	MVGKMKKITAMPKMLIG--AKGCTHTFTVTYQVTHMSGGAIMGEDPKTSAVNRYLQ	563
Qy	487	VWDFANLYVAGNGTIRTFGENPT	510
Db	564	SWDVPNVFVPGASAFPOGLGYNPT	587
RESULT 6			
Q9HQH8	IC	Q9HQH8 PRELIMINARY; PRT; 529 AA.	
AD	Q9HQH8		
DT	01-MAR-2001 (TEMBLrel. 16, Created)		
DT	01-MAR-2001 (TEMBLrel. 16, Last sequence update)		
DT	01-DEC-2001 (TEMBLrel. 19, Last annotation update)		
DN	VNG1035C.		
GN	VNG1035C.		
OS	Halobacterium sp. (strain NRC-1).		
OC	Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;		
OX	Halobacterium.		
OC	NCBI_TaxID=64091;		
RN	[1]		
RX	SEQUENCE FROM N.A.		
RP	MEDLINE=20504483; PubMed=11016950;		
RA	Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,		
RA	Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sirogna J.,		
RA	Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,		
RA	Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,		
RA	Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,		
RA	Isenbarger T.A., Peck R.F., Ponischroder M., Spudich J.L., Jung K.-H.,		
RA	Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,		
RT	Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;		
RT	"Genome sequence of Halobacterium species NRC-1."		
RL	Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).		
DR	EMBL; AF005037; RAG19445.1; -		
DR	InterPro; IPR000425; MIP.		
DR	InterPro; IPR000205; NAD_binding.		
DR	PROSITE; PS00221; MIP; UNKNOWN_1.		
SK	Complete proteome.		
Q9	SEQUENCE 529 AA. 56633 MW; FAED92051336225D CRC64;		

```

Query Match          5.5%  Score 158.5;  DB 17;  Length 529;
Best Local Similarity 22.3%;  Pred. No. 7.2e-05;
Matches 117;  Conservative 44;  Mismatches 194;  Indels 169;  Gaps 22;

QY      84  A Y R V G V G M S T H T C S T P R T H P P --- M E S L P G I G R P K L S N D P A D D K E W N E L Y S E A E R L I 140
      | : : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db      88  A R V K G V G S T L H W G M V M R L H E Q D F R I A S A T G V G A --- D W P E D Y D T L K P Y Y A A A E S A L 142

QY      141  -- G T S T K E D E S I R H T L V L R S L Q A V K D R Q I R F R P L P L A C H R L K N A P E Y V E W H S A E N L F 197
      | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db      143  G V S G A S N P F A P P R E Q P H Q P A F P P S T S D --- S L F A D --- A C E S L G I A T --- 185

QY      198  H S I Y N D D K O K K L F L L N H R C T R I A L T G --- G Y E --- K K I G 232

```

Db	186	HSVNP	-----ARLSAGRETRRACVGYGTQPCVPCPSGAKYDATVHVDRATDAGARVID	237
Qy	233	AAEVRNL	-----LATRNPSSQLDSYIMAKVYVVLASCAIGNPQIL	272
Db	238	EAPVORLEDAAGDRVTCVAVTADGTTHRQS	---ATEFVLAAGGIETPRLLLSDSRY	294
Qy	273	NSGF	---SGLQVTPRNDSLIPNLGRIETQPMAPFCVLIRQEFVDSVRDDPYGLPWNKEA	329
Db	295	PDGLANSSGL	-----VGRY-----FMD-----	311
Qy	330	VAQIAKNPTDALPIPRDPEPQVTTFTTEHPMHMTOIHRDAFSYGAV	-----GPEV	381
Db	312	---HLFAGAGTLDPTQRNHVGENT	---TESHQY---DRPDGSRGAIKLEFLNYAGPSP	363
Qy	392	DSRIVDLRWFGA	-----TDPEANNLVFNQDQDGYSPQPOTFR	421
Db	364	AEMALSGDDGDGDCMDRIRDASGTHIAVGLVEQQPRPNRVLHPERTDVHGNPVPDVV	423	
Qy	422	YRPSTASNVARKMMADMCVEASNLGGYLP	-TSPQPMDPCLALHLAGTTRIGFDKATTV	480
Db	424	WLSAYERRTIERANEIOREILLTGADIEWTVGPE	-DTGPAFHMGTTTRMGCTDPAESV	481
Qy	481	ADNLSLVMDFANLYVAONGTIRTGFGENPNTITSMCHAIKARS	524	
Db	482	VDPLRTHDLSNLVSSSVFPFTAGAMNPTLTIAALKAADHI	525	

RESULT

7

Q9LBK8

ID

Q9LBK8

PRELIMINARY;

PRT;

545 AA.

DT

01-OCT-2000 (TReMBLrel. 15, Created)

DT

01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DT

01-OCT-2000 (TReMBLrel. 18, Last annotation update)

DE

SOLBITOL DEHYDROGENASE LARGE SUBUNIT.

GN

SLDL.

OS

Glucanobacter oxydans (Glucanobacter suboxydans).

OC

Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;

OC

Glucanobacter.

OX

NCBI_TaxID=442;

RN

[1]

RP

SEQUENCE FROM N.A.

RC

STRAIN-IFO 3254;

RA

Yoshikawa K., Saito Y., Ishii Y., Noguchi Y., Soeda S.;

RL

"PCT Patent Publication No. WO9920763A1.";

RT

Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RE

EMBL; AB039821; BAA92378.1; -

DR

InterPro; IPR000205; NAD_binding.

SQ

SEQUENCE 545 AA; 60076 MW; 4C0A4F3C56950E22 CRC64;

Query Match

5.4%;

Score 157;

DB 2;

Length 545;

Best Local Similarity

19.7%;

Pred. No. 0.0001;

Matches 116;

Conservative 69;

Mismatches 198;

Indels 206;

Gaps 25;

Qy	34	KGALQOVSVPRNQNVPTLDPGAWSPGSSAISNG	-----KNPHREFENLSAEAVTRGV	89
Db	58	KGAYQLPYPPV	-----FWAMHPDQGPNGYLHTTGPDGAAY	102
Qy	90	GGMSTHTWCSTPRHP	--PMESI.PGIGRPKLSNDPAEDDKENNELYSAEALIGTSTKE	146
Db	103	GGTTHWAGCANRYLPSPDFELHSRYGVR	-----DWAIKYDDLEPPYQAEVMNGVAGPN	157
Qy	147	FDESIRHTLVLRSLQDAYKORQIF	--RPLFLACHRLKNAPEVVEHSAENLFHSIYND	204
Db	158	MD	-----VDDLGSFSPRSHNYPMKVEPLS	183
Qy	205	KQKLFLLTNHR	-----CTRLALTGG	231
Db	184	QFRKLIIHEKNTYRNVHPVQPARNTPYDKRPTCEGNNCMPICPIGAMYNGIHVSNNHAEAA	243	
Qy	232	GAABVRNLLATRNPSSOLDSYM	-----AKVYVLAGSAI	265

Query Match	5.4%;	Score 157;	DB 2;	Length 545;
Best Local Similarity	19.7%;	Pred. No. 0.0001;		
Matches 116;	Conservative 69;	Mismatches 198;	Indels 206;	Gaps 25;
Qy	34	KGALQQVSPVRNQNVPTLDPGAWSAAPGGSSAISNG	-----KNPQREFENLSAEATRGV	89
Db	58	KGAYQLPYPPV	-----PWAMHPPDQSPGVLHTTGDGAYQ	102
Qy	90	GGMTHTWCTSTPRIHP	---PWESLPGIGRPKLSNDPAEDDKWENLYSBAERLIGTSTKE	146
Db	103	GGTTWHWAGCAWRYLPSPDFELHSRYGVGR	-----DWAIKYDDLEPFYQAEVNMGVAGPN	157
Qy	147	FDESIRHTLVLSQDAYKQORIF	--RPLPLACHRLKNAPEVWEHSAENLFHSLYNDD	204
Db	158	MD	-----VDLLGSPRSNHPMKVEPLS	183
Qy	205	KQKLFLLTNHR	-----CPTRLALTGG	231
Db	184	QFRKLIHEKATNYRVVHEPQARNTRPYDKRPTCEGNNCMPICPIGAMYNGIHSVNAEAA		243
Qy	232	GAARVRLNLTATRPSSOLDYSIM	-----AKVYVLASGAI	265

```
Db 244 GARIPNAVVRLETDASNKVVPVNYDDPKNSHRVTGFFVVAACIESAKLLLLLSAD 303
QY 266 -GNPQILYNSG-----FSLQVT--PRNDSLIPNLGRIYITEQPMFACQIVLRQEFV 313
Db 304 DKNPRGIANSSDQVGRNMMDHTGVQLFSMGNSDSLWPGRGPLLT-----SII 350
QY 314 DSVRDDYGLIPWKE--AVAQHIAKNPTDALPIPRDEPQVTPPTFEHPWHHTQIHRDA 371
Db 351 DSFRDG-----PWRSERGAYLVHMYD-----DNOVDFAATGL-----A 382
QY 372 FSYCAVGPVDSRVID-----LRWF-----GATDPEANLLVFQNDVDGYSMPOPTFRYR 423
Db 383 IAKGVCKELEEQIRYGSNAVRULFSNENGIADPD--NRUTLSKTHKDVIGIPIPEVYIK 440
QY 424 ---PSTASNRARKMADMCCEVASNLG-----GYLPTSPQFMGPGGLALHLAGTTRIGFD 475
Db 441 LPEYTVKSCDHTKELFELMALMSGTDPOWTKGYFP-----QCHPSGSTINGTD 489
QY 476 KATTVADNNSLVDFANLYVAGNTITRTGFGENTPLTSMCHAIKSARSI 524
Db 490 PTNSVVDGECRTHDENLFARSASFSSVGTGNTITLIGALALRAAASL 538
```

```
RESULT 8
Q9XCR0 PRELIMINARY; PRT; 553 AA.
AC Q9XCR0;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE 2-KETO-GLUCONATE DEHYDROGENASE SUBUNIT.
GN KDBG.
OS Pantoea citrea.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=53336;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20200361; PubMed=10735866;
RA PuJol C.J., Kado C.I.;
RT Genetic and Biochemical characterization of the pathway in Pantoea
RT citrea leading to pink disease of pineapple.;
RL J. Bacteriol. 182:2230-2237(2000).
DR EMBL; AF131202; AAD4706.1;
DR InterPro; IPR000205; NAD_binding.
SQ SEQUENCE 553 AA; 59939 MW; F2A60424416B8C3F CRC64;
```

```
Query Match 5.1%; Score 148; DB 2; Length 553;
Best Local Similarity 20.6%; Pred. No. 0.00062;
Matches 114; Conservative 59; Mismatches 222; Indels 160; Gaps 21;

QY 52 LDPCANAPGSSAISGNKPNHQREFNLSAEATRVGGMSTHTWTSTPRIHPP---ME 108
Db 81 LAPAPLYFPNNY--NVTGPSAGSFQ---QGYLRTVGGTTHWAASCRHHPSDFVMK 134
QY 109 SLPGIGRP-KLSNDPAEDDKWENLYSEAERLIGT-----STKEFDESIRHTLVLR 158
Db 135 SKYGVGRDWFISYDEME---PW---YCAEYEGVAGPSDMSQSPSERSRPYMDWVPF 188
QY 159 SLQDAY-----KDRQIRFP---LPLAC-----HRLKN 183
Db 189 AHGDTYFASVVPNHYNLVPIQGRSTRPWEGRPVCCGNNNCPICPIGAMNGIHIER 248
QY 184 APEVVEHSAENLPHSYNDOKKLLTLLTNHRCR---LALVGGYEKKIGAAEVNLL 240
Db 249 AESKGAVYLAESVYVYKIDTD-----NNRVTAHVHLDNQGASHKATG----- 290
QY 241 ATRNPSSQLDSYIMAKVYVYLASGAIGNPQIILYNSGFSGLQVTPRNDSLIPN---LGR-- 294
Db 291 -----KAFALACNGIETPRLLQA-----ANKNPTGIANSMDWGRNM 329
```

```
QY 295 -----YITEQPMFACQIVLRQEFVDSVRDDPYGLPWWKEVAQHIANKPTDALPIP 345
Db 330 MDHSGFHCSELTPEPVWLGRRGPAQSSCMVGPDRDGAFRSEY---SANKMILNNISRVVPAT 386
QY 346 FRDPEPQVTPPTFEHPWHHTQIHRDAFSYAGVGPVDSRVI-----VDLRWFGATDEA 399
Db 387 -----KOALAKGLVGRKALDEEIRYSIHGYDLSISLEPLDP 423
QY 400 NNLVLFQNDVDGYSMPOPTFRYRPESTASNRARKMADMCCEVASNLGGLYLTSPPOQMD 459
Db 424 ENRUTLSKTRKDPGLACPDHYDGYVRKGATAAHEQLOHIGSLFNG-----KEFN 476
QY 460 PGLAL----HLAGTTRIGFDKATTVADNNSLVDFANLYVAGNTITRTGFGENTPLTSMC 515
Db 477 ITTALNANNHMGITMGSKAKDAVDGNCRTFDHENLWLPGGGAIPASVYVNSTLSMAA 536
QY 516 HAIKSARSINTLK 529
Db 537 LGLKAAHDI SLRMK 550

RESULT 9
Q9RZ26 PRELIMINARY; PRT; 722 AA.
AC Q9RZ26;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE GMC OXIDOREDUCTASE.
GN DRA0127.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RI;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RT Science 286:1571-1577(1999).
DR EMBL; AE001862; AAFI2230.1;
DR TIGR; DRA0127;
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR001199; Cyt_B5.
DR InterPro; IPR000172; GMC_oxred.
DR InterPro; IPR001899; Gram_pos_anchor.
DR Pfam; PF00732; GMC_oxred; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR PROSITE; PS00191; CYTOCHROME_B5_1; UNKNOWN_1.
DR PROSITE; PS00624; GMC_OXRED_2; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 722 AA; 75376 MW; 450DF1CD1B7596F9 CRC64;
```

```
Query Match 5.0%; Score 144; DB 16; Length 722;
Best Local Similarity 21.1%; Pred. No. 0.0021;
Matches 107; Conservative 68; Mismatches 196; Indels 136; Gaps 26;

QY 90 GGMSTHTCTSPRIHPMESLPGICRPKLSN-----DPAEDDKWENLYSEAERLIG 141
Db 277 GGSVNNNSV-----PPRDDI----RQWASEHGLSDVADPGYD----RHDAVLERM-G 323
QY 142 TSTKEFDESIRHTLVLRSLQDAYKDRQIRFPRLPLACHRLKNAPEYVEHSAENLPHSY 201
```

Db 324 VSEQSDHNGPHQ---RLVEGADKLGTYTVKA-----ALNLSPEH---YDADKAGHAGF 371
 QY 202 NDD---KOKKLEF-----LITNHRCTRLALTGGYEKKIGAAEVRNLLATRNPPSS 247
 Db 372 GQGTGAKGCTLNTFLKDAFEAGARILLVGTQRAQRLVEDG-----RAAGVSAVTTMGDETR 426
 QY 248 QLDSTYMAKVVYVLAIGAIGNPOLYNSGFGSLQVTPRNDSLIPNLGRYITEQPMAPCQIV 307
 Db 427 QIT--VRAPQVVVACGAETPALRLSGIGG-----PAAGRYLRLHPAGIVAGI 473
 QY 308 LRQEFVDSVRDDPYG--LPWKEAQAQIAKNPTDPIPERDPE-----P 351
 Db 474 -----YGEDQRAWGP-----POSGILKQFADHENGHGFIEGVQVQYGP 511
 QY 352 QVTT---PPT---EEH--PHTQIHRDAFSYGAVGPEVDSRIYVD-----LRWFGATDP--EA 399
 Db 512 ALMASGLPWTGCEAHRDLMSKFRHWATFVSIYQDRGHGQVTVDDGNAVHTYALTDLLDA 571
 QY 400 NLLVFNQDVQGYSM-----POPTFRYRPSTASNVARKMMADCEVAS-----NLGGYLP 451
 Db 572 RN---FRRGVTESIRLHEAAGAEIEVALAPGVAPWRRGDDLEAFIGQVAQVPLGAGQTV 628
 QY 452 TSPQPQMDPGLHLAGTTRIGFDKATTVADNNSLVWDFANLYVAGNCTIRTFGCGENPIL 511
 Db 629 FS-----AHQMSARMGSDQTSVADPDGQLHDVPGVWIGDTSAPFTCSGVNPMV 678
 QY 512 TSMCHAIKARSIIINTLKGTDGKNTG 538
 Db 679 SCMALASRTAEKLLAAMEGADGTGSG 705

RESULT 10

Q98C76 PRELIMINARY; PRT; 523 AA.
 AC Q98C76;
 DT 01-OCT-2001 (TRENBLrel. 18, Created)
 DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
 DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)
 DE MLR5266 PROTEIN.
 GN MLR5266.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.,
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP003006; BAB51745.1; --
 KW Complete proteome.
 SQ SEQUENCE 523 AA; 58536 MW; F026E42A6F9E3BE8 CRC64;

Query Match 4.9%; Score 141; DB 16; Length 523;
 Best Local Similarity 20.9%; Pred. No. 0.0022;
 Matches 101; Conservative 60; Mismatches 194; Indels 128; Gaps 20;
 QY 86 TRGVCGMTHWTCSTPRHP---PMESLPGIGRPLKND---PAEDDKENNELYSERLI 140
 Db 124 TIGLGGGNCNFAQTPFRHPNDRFLKSTYIGG-----NDWPISYDDVE--PPYCDAAEII 176
 QY 141 GTS-----TKFDESIRHLTVLSLQDAYKDRQIRFP-----LPLA-----CH 179
 Db 177 SISGDPDMMMLPRSKPFPQPHRMSTPDKMKMAAQPDQHFVMPARARVPTAQTSCCA 236

QY 180 RLK--NAPYEVWHSAENLFHSIYNDKOKKLFLLTINHRCTRLALTGGYEKKIGAAEYR 237
 Db 237 NLRCWLCPPVDAKFTANGLMHVFEHPD-----VSVCLGAEVRRLDQVGG-----TVR 283
 QY 238 NLLATRNSSQLDSYIMAKVYVYVLAIGAIGNPOLYNSGFGSLQVTPRNDSLIPNLGRYIT 297
 Db 284 SVTFVHDGK---EYQVSGDLFILGANAIQSAAILRSG----- 318
 QY 298 EQPMAPCQIVLRQEFVDSVRDDPYGLPW---WKEAQAQIAKNPTDPIPERDPEQV 353
 Db 319 -----LRDEFVGRGLHESYG--WNFEVYLDGVDFDGTSTTTGLNFGLYD----- 361
 QY 354 TTPFTTEHPWHTQIHRDAFSYGAVGPEVDSRIYVDLRFMGATDPEANNL-----LVFQ 406
 Db 362 -GPHRSEH-----AAALVYFENRWQHGMRKGLRQLTLPVVVTE 401
 QY 407 N--DVQDGYSMQPTFRYRPSTASNVARKMMADCEVASNLGGYLPSPQPMDFGL-- 462
 Db 402 NLLDDENFVTLDEDDNAFVSEKAPSDYAVKGMARALDKLPPELLAPLVE--RLFDGRIRP 459
 QY 463 -ALHLAGTTRIGFDKATTVADNNSLVWDFANLYVAGNCTIRTFGCGENPILTSMCHAIKSA 521
 Db 460 TESHVQGLRWGTGADSVIDSNNIHLRLNLVVVGTSTYTPSCSCANPELTAALSLRAA 519
 QY 522 RSI 524
 Db 520 SRI 522
 RESULT 11
 Q911K8 PRELIMINARY; PRT; 591 AA.
 AC Q911K8;
 DT 01-MAR-2001 (TRENBLrel. 16, Created)
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
 DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)
 DE GLUCONATE DEHYDROGENASE.
 GN PA2265.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=2043737; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.D., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 DR EMBL; AE004652; AAG05653.1; --
 KW Complete proteome.
 SQ SEQUENCE 591 AA; 64716 MW; 9A7012B67ECEB9DE CRC64;

Query Match 4.9%; Score 141; DB 16; Length 591;
 Best Local Similarity 19.9%; Pred. No. 0.0027;
 Matches 120; Conservative 63; Mismatches 235; Indels 184; Gaps 22;
 QY 45 RNQVPTLDPGAW-----SAPPOSSAISNCKNPHQREFEMLSAEAVTR 87
 Db 38 RGENRDYTPDGAYPNTLDYNTYRGLFQNLKSTVSIIRGINDTALPYRLSALFPGD 97
 QY 88 GVGGMSTWTCSTPRHPMESLPG-----IGRPKLSNDPAEDDKENNELYSER 137
 Db 98 GVGAGLHWSGVHFRIMPBEELRLSHRYEERYGKKFIPGMITQDYGVSVEELPHDFPAE 157
 QY 138 RLIGTSTKEFDESIRHLL-----VLRSIQDAYKDRQIRF----- 172


```

Db 158 KVFQTS-----GTAHTVKGQVVGKGNPFPAADRSDDPPLPALRQVYS--AQLERKAAEEL 209
QY 173 -----PLPLA-----CHRLKNAPEYWEHSAENLFHSIYNDKQKK 208
Db 210 GLHPYDLPAANASGPWNPYGVOMGFCNFCGFCGYACYWYKASPNL--NILPALRQTP 267
QY 209 LFTLLNHRCTRRLALGGYEKKIGA-----AEVRNLLATRNPSOLDOSYIMAKYVYVLASGA 264
Db 268 LFEELRANCNVLKVNLDSDGRQATGVTVYDAQGREIVQP-----AKLVIISAFQ 315
QY 265 IGNPQILYNSGSGSLQVTPNDSILNLRXITEQPMACQIVLRQEFVDSVRDPY--- 321
Db 316 FHNVRLLLSGI-GKPYDPRTGEGV--VGKNFAYQNNM-----TIKAFDQKDVHTNPFVGT 368
QY 322 -----GLPWWKEAVAQHIKAKNPTDALPIPRDPPEQVTTTF 357
Db 369 GGGGVAVDNADNHDGPLGFGVGGSPW---VNQAGSKPIGGLAVPPCTPS----- 417
QY 358 TEHPHPT---QTRDAFSYGAVGPEVDSRVIVDLRFGATDPEANNLLVFQNDVQDYS 414
Db 418 -----WGSQKQAVKDAYTH-----TVSMDAHGSMNTYRNYLDLDPYKDAYG 461
QY 415 MP-OPTERY-----RPTASNVARKMMADMCEVA-----SNLGGYLPITSPPOFMDPGL 462
Db 462 QPLLRTFTDKNEIRMSRVVTEHMRKIAEAMPKALSVSKNFGDHFNTRVYQ----- 515
QY 463 ALHLAGTTRIGFDKATTADVNNSLVWDFANLYVAGNGTIRTFGGENPTLTSCHKAHSAR 522
Db 516 TTHLGGGAINSGDPKTSVLNRYLQSDVHNVFVNGASAFPGQGYNPPTGLVAALAYWSAK 575
QY 523 SI 524
Db 576 AI 577

RESULT 12
Q9M0H4 PRELIMINARY; PRT; 748 AA.
AC Q9M0H4
AT Q9M0H4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE HYPOTHETICAL 82.0 KDA PROTEIN.
GN AT4G28570.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Lennard N., Quail M., Harris B., Rajandream M.A., Barrell B.G.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL161573; CAB81445.1;
DR HSSP; P22637; 3COX.
DR InterPro; IPR000205; NAD_binding.
KW Hypothetical protein.
SQ SEQUENCE 748 AA; 81959 MW; 13CD183F5940DAC3 CRC64;

```

```

Query Match 4.4%; Score 128.5; DB 10; Length 748;
Best Local Similarity 19.7%; Pred. No. 0.046;
Matches 105; Conservative 77; Mismatches 187; Indels 165; Gaps 24;

QY 76 EFNLSAEAVTRGVGGMSTHTCTSTPRIHPMESLPGIGRPKLSNDPAEDDKENNELYSE 135
Db 301 KFMLLAGSAGV---GGTAVNWSAS---IRTPDHVL-----QEM-----SE 334

```

```

QY 136 AERLICHTSTKEF-----DE-SIRHTLVLRSLQDAYKDRQIRFRPLPLACHRL-----K 182
Db 335 GSKIKFSGSEQYQAMDEVIRIGVTERCVKHGFQN-----QVLRKGCERGLQVESVPR 389
QY 183 NAFB-----YVEHSAENLFHSIYNDKQKKLFTLLNHRCTRRLAL-----TGGYEKK 230
Db 390 NSPDHYCGLCGCGCRAGAKNGTDQTLVDAVENGAVILTGKAERFVLVDNTSSSSNERK 449
QY 231 -----IGAAEVRNLLATRNPSOLDOSYIMAKYVYVLASGAIGNPOILYNSGSGSLQVTPRN 285
Db 450 KRCVGFPASSVGKIGKK-----FIIEARVTVSSAGSLTTPMLSLSSGLKN----- 495
QY 286 DSLIPNLGR-----YITEQPMACQIVLRQEFVDSVR--DDPYGLPWWKEAVA 331
Db 496 ----PNIGRNLKHPVLMTGWYFPEKDSFGKMEGGIITSVHMND-----TESGC 544
QY 332 OHTAKNPTDALPIPRDPPEQVTTPTTEHPHPTQIHRDAFSYGAVGPEVDSRVIVDLRW 391
Db 545 KALENPLIG-PASYAGLSPWVS-----GPDLEKRMK----- 576
QY 392 FGATDPEANNLLVFQNDVQDYSM--PQPTFRYRPSSTASNVARKMMADMCEVAS---NL 446
Db 577 YGRT-----AHLFALVRDLGSGEVMENEVYRTTKDRENLRAGLRQALRVSVAAGAVEV 632
QY 447 GGY-----LPTSPPOFMDPGLAL-----HLAGTTRIGFDKAT 478
Db 633 GTVRSDDQMKCEAITKEAMEEFLDEVDVAGVGVTGKEYTTFSAHQMSGRMGVTAAE 692
QY 479 TVADNLSLWDFANLYVAGNGTIRTFGGENPTLTSCHKAHSARSIINTLKGGT 532
Db 693 GALDENGESWEAEGLFVCDGSLPSAVGVNPMITIQSTAYCISSKIIVDSLQNK 746

RESULT 13
Q94BP3 PRELIMINARY; PRT; 748 AA.
AC Q94BP3;
AT Q94BP3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE HYPOTHETICAL 82.0 KDA PROTEIN.
GN AT4G28570.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Dale J.M., Gibson H.A., Goldsmith A.D., Jiang P.X., Lee J.M.,
RA Quach H.L., Tang C., Toriumi M., Yu G., Bowser L., Carinci P.,
RA Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full Length cDNA of gene AT4G28570 (GI:7269712).";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY039977; AAK64154.1;
KW Hypothetical protein.
SQ SEQUENCE 748 AA; 81987 MW; E3EB613F5D56A0D1 CRC64;

```

```

Query Match 4.4%; Score 128.5; DB 10; Length 748;
Best Local Similarity 19.7%; Pred. No. 0.046;
Matches 105; Conservative 77; Mismatches 187; Indels 165; Gaps 24;

QY 76 EFNLSAEAVTRGVGGMSTHTCTSTPRIHPMESLPGIGRPKLSNDPAEDDKENNELYSE 135
Db 301 KFMLLAGSAGV---GGTAVNWSAS---IRTPDHVL-----QEM-----SE 334

```

QY 136 AERLIGSTKPE-----DE-SIRHTLVLSQDAYKDRIFRPLPLACHRL-----K 182
 Db 335 GSKIRFGSQEYQSAMDYITIRIGTVGTCVKGFGN-----QVLRKGCERLQLQVESVPR 389
 QY 183 NAPE-----YVEHSAENLFHSIYNDKOKKFLTLTNRHCRTRLAL-----TGGYEKK 230
 Db 390 NSPEDHYCGLGCGRAGAKNKTDTQTLVDVAVENGAVILTGKAERFVLVDNTSSNERK 449
 QY 231 -----IGAAEVRNLLATRNPPSSQLDSYIMAKYVVLASGAIGNPQILYNSGSLQVTPRN 285
 Db 450 KRCVGFVASSVGKIGKK-----FIIEARVTVSSAGSLTTPPLMLSSGLKN-----495
 QY 286 DSLINLGR-----YITQPMFACOIVLRQEFVDSVR--DDPYGLPWNKEAVA 331
 Db 496 -----PNIGRNLKHPVMTGWTFPEKDEFSKGMTEGGIITSVHMND-----TESGC 544
 QY 332 OHIAKNPTDALPIPRDPEPQVTTFTTEHPWHTQIHRDAFSYGAGVEPDSRVIVDLRW 391
 Db 545 KAILLENPLIG-PASVAGLSPWVS-----GPDKERMIR-----576
 QY 392 FGATDPEANLLVFQNDVDGYSM--POPTFRYRPSTASNVARKMADMCVAS---NL 446
 Db 577 YGRT-----AHLFALVRDLGSGEVMMENEVYTTKKDRNLRLAGLRQALRVSVAGAVEV 532
 QY 447 GGY-----LPTSPPOFMDPGLAL-----HLGATTRIGFDKAT 478
 Db 633 GTYRSDGQMKCEAITKEAMEEFLDEVDAVGGVGTGKEYTWTYFSAHQWCSRMGVTAEE 692
 QY 479 TVADNNSLWMDFANLYVAGNCTIRTFGENPTLTSMCHAISARSIIINTLGGT 532
 Db 693 GALDENGESWEAEGLFVDCGSLPSAVGVNPNMTITQSTAYCISSKIVDSLQNK 746

RESULT 14

Q988P2 PRELIMINARY; PRT; 499 AA.
 AC Q988P2;
 DT 01-OCT-2001 (Tremblrel. 18, Created)
 DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE MLR6655 PROTEIN.
 GN MLR6655.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 DR EMBL: AP003009; BAB52905.1; -
 DR InterPro: IPR000205; NAD_binding.
 KW Complete proteome.
 SQ SEQUENCE 499 AA; 54537 MW; 633924AED9688F51 CRC64;

Query Match 4.4%; Score 127; DB 16; Length 499;
 Best Local Similarity 20.68; Pred. No. 0.033;
 Matches 118; Conservative 58; Mismatches 223; Indels 174; Gaps 26;

QY 13 GYHKNEIEFQKDRFVNIKGLQQLQYVPRNQNVPTLDPGAWSPGSSAI-----66
 Db 37 GEHLRDTPEARDIAIFQNGVRSSE-----WLAIDGESFLPGNYYY 79
 QY 67 --SNGK-----NPHOREFENLSAEAVTRCVGMSTHWCSTPRIHP---PM 107

Db 80 VGGNSKFFGAMVRYRQEDFNPRDH-----MGRSPGWPISYAELEPWYERA 126
 QY 108 ESLPGIGPKLSNDPAEDDKENNELYSEARLIGTSTKEFDESIRHTLVLSLQDAYKDR 167
 Db 127 ELLFGV-RGDARQDTEPPR--NRPY-----RYLPVDPDEPAIATVQRRLQAGIHFA----175
 QY 168 QRIFRPLPLA-----CHRLKNAPEYVEVHSAENLFHSIYNDKOKKFLTLTNRHCRTRL 221
 Db 176 -----SLPLAIDIDAWLRRAKTG-----WDAPPNT--GAGKIDAEGVPLTKALEHPNATL 223
 QY 222 ALTGGEYKKGKGAEEVRNLLATRNPPSSQLDSYIMAKYVVLASGAIGNPQILYNSGSLQV 281
 Db 224 ITGANVQRVLVTDAAGRRVMAAVFIKDGVELSIGADVFAVAGVQSAALLLRS-----276
 QY 282 IPRNDSLIPN-----LGR-YITEQPMFACQI-----VLRQE---FVDSVRDDPYG 322
 Db 277 ---STSVYPNGLGNSSDQIGRNFNMHNTAMLAIDPFRNRNTAVYQKTLGDFNDFYKNKPLG 333
 QY 323 -LPWNKEAVAQHIKAKNPTDA---LPIPRDPPEQVTTFTTEHPWHT-QIHRDAFSYGA 376
 Db 334 SFPLGNVOLLGHITGNILKANAPLLP-----RWLAGLVARNICYGWFL 375
 QY 377 VG---PEVDSRVIVDLRWFGATDPEANLLVFQNDVDGYSMPOPTFRYRPSTASNVRAR 433
 Db 376 TSEDLPNPESRVTI-----RNGRIVMNVVRNMGAHETLIR-----RTR 414
 QY 434 KMW--ADMCEVASNLGGYLPSPPOFMDPGLALHLAGTTRIGFDKATTVADNNSLWMDFA 491
 Db 415 AVMRAGEFPVVLTRTFGRKTTs-----HOCGARLGSDPNTSVVSPDCRSHDIA 463
 QY 492 NLYVAGNCTIRTFGENPTLTSMCHAISARSII 524
 Db 464 NLYVTDSVLTSRAVNPALTIALAKAGRAI 496

RESULT 15

Q12623 PRELIMINARY; PRT; 620 AA.
 AC Q12623;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE GLUCOAMYLASE (EC 3.2.1.3).
 GN GLA1.
 OS Humicola grisea.
 OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
 OX NCBI_TaxID=5527;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Berka R.M., Rey M.W., Thompson S.A., Gray G.L., Carmona C.L.,
 RA Power S.D.;
 RT "Molecular cloning, analysis and expression of the glal gene encoding
 a thermostable, raw starch-digesting glucoamylase from the fungus
 Humicola grisea var. thermoides.";
 RT Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.
 RL EMBL: M89475; AAA33386.1; -
 DR HSSP: P04064; ICAI
 DR InterPro: IPR002044; CBD_4.
 DR InterPro: IPR000165; Glyco_hydro_15.
 DR Pfam: PF00686; CBD_4; 1.
 DR Pfam: PF00723; Glyco_hydro_15; 1.
 DR PRINTS: PR00736; GLHYDRASE15.
 DR ProDom: PD001568; CBD_4; 1.
 DR PROSITE: PS00820; GLUCOAMYLASE; 1.
 KW Hydrolyase; Glycosidase.
 SQ SEQUENCE 620 AA; 66525 MW; 54214FF67E20BE0A CRC64;

Query Match 4.1%; Score 117.5; DB 3; Length 620;
 Best Local Similarity 19.08; Pred. No. 0.3; Mismatches 94; Indels 189; Gaps 30;
 Matches 124; Conservative 94;

This Page Blank (uspto)